```
              GenCore version 5.1.6
         Copyright (c) 1993 - 2003  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:   December  1, 2003, 07:30:31 ; Search time 2775 Seconds
                           (without alignments)
                           10045.850 Million cell updates/sec

Title:          US-10-049-742-22
Perfect score:  1147
Sequence:       1 gccttgggtcaagcagaata........gcctcctgagtagctgggat 1147
Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       22781392 seqs, 1215238056 residues

Total number of hits satisfying chosen parameters:      45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   EST:*
          1: em_estba:*
          2: em_esthum:*
          3: em_estin:*
          4: em_estmu:*
          5: em_estov:*
          6: em_estpl:*
          7: em_estro:*
          8: em_htc:*
          9: gb_est1:*
         10: gb_est2:*
         11: gb_htc:*
         12: gb_est3:*
         13: gb_est4:*
         14: gb_est5:*
         15: em_estfun:*
         16: em_estom:*
         17: em_gss_hum:*
         18: em_gss_inv:*
         19: em_gss_pln:*
         20: em_gss_vrt:*
         21: em_gss_fun:*
         22: em_gss_mam:*
         23: em_gss_mus:*
         24: em_gss_pro:*
         25: em_gss_rod:*
         26: em_gss_phg:*
         27: em_gss_vrl:*
         28: gb_gssl:*
         29: gb_sss2:*
```

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	895.2	78.0	942	13	BU538214	BU538214 AGENCOURT
2	796.8	69.5	2300	11	AK017830	AK017830 Mus muscu
3	796.8	69.5	2712	11	AK045445	AK045445 Mus muscu
4	795.2	69.3	2639	11	AK032681	AK032681 Mus muscu
5	795.2	69.3	4231	11	AK082579	AK082579 Mus muscu
6	792.6	69.1	918	13	BU956189	BU956189 AGENCOURT
7	787.4	68.6	3022	11	AK045345	AK045345 Mus muscu
8	786.4	68.6	4053	11	AK083878	AK083878 Mus muscu
9	764.6	66.7	957	12	BM802135	BM802135 AGENCOURT
10	759.4	66.2	878	12	BG763500	BG763500 AGENCOURT
11	744.2	64.9	955	13	BU538351	BU538351 AGENCOURT
12	706.4	61.6	785	10	BI333110	BI333110 60296387
13	692.8	60.4	904	13	BQ691177	BQ691177 60296387
14	672.2	58.6	995	12	BM452174	BM452174 AGENCOURT
15	644.2	56.2	723	10	BE619407	BE619407 601473229
16	639	55.7	906	10	BG676002	BG676002 60222366
17	622	54.2	1046	12	BM802913	BM802913 AGENCOURT
18	605	52.7	605	13	BX090711	BX090711 BX090711
19	603.6	52.6	723	10	BE779250	BE779250 601464514
20	601.8	52.5	940	10	BG328197	BG328197 602427378
21	580	50.6	874	10	BF679562	BF679562 602153994
22	573.8	50.0	927	12	BM424147	BM424147 AGENCOURT
23	566	49.3	588	9	AA460131	AA460131 zx68h04.r
24	565	49.3	1046	12	c BQ058270	BQ058270 AGENCOURT
25	562.2	49.0	912	10	BG751088	BG751088 602729879
26	559	48.7	737	14	CB324690	CB324690 UI-R-D20-
27	550	48.0	618	12	BM011379	BM011379 603635574
28	548.6	47.8	805	10	BE536083	BE536083 601061730
29	546.8	47.7	864	10	BG252385	BG252385 602365980
30	541.4	47.2	679	9	c AW146163	AW146163 um38b12.x
31	533.4	46.5	670	12	BU708368	BU708368 UI-M-FC0-
32	530.4	46.2	667	14	BY760298	BY760298 BY760298
33	524.2	45.7	756	12	BG868350	BG868350 602787804
34	520.4	45.4	900	10	BG298053	BG298053 602396271
35	494	43.1	776	12	BG821631	BG821631 602727634
36	490.4	42.8	651	10	BB465186	BB465186 BB465186
37	486.2	42.4	871	10	BF139916	BF139916 601789432
38	478	41.7	560	9	c AV602887	AV602887 AV602887
39	467	40.7	469	10	BG504969	BG504969 602551409
40	465.2	40.6	527	12	c BI002288	BI002288 PM1-HN007
41	458.8	40.0	598	12	c BM208252	BM208252 C624E11-
42	436.6	38.1	448	12	BI002295	BI002295 PM1-HN007
43	424.6	37.0	781	14	CB520550	CB520550 UI-M-GI0-
44	419.4	36.6	505	10	c BF819630	BF819630 MR1-RT002
45	405	35.3	968	13	BU838931	BU838931 AGENCOURT

ALIGNMENTS

```
RESULT 1
BU538214
LOCUS       BU538214              942 bp    mRNA     linear   EST 13-SEP-2002
DEFINITION  AGENCOURT_10186479 NIH_MGC_107 Homo sapiens cDNA clone
            IMAGE:6658738 5', mRNA sequence.
ACCESSION   BU538214
VERSION     BU538214.1  GI:22848655
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 942)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2754  row: o  column: 10
            High quality sequence stop: 680.
```

```
FEATURES             Location/Qualifiers
    source           1..942
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6568738"
                     /tissue_type="adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_107"
                     /note="Organ: breast; Vector: pOTB7; Site_1: EcoRI;
                     Site_2: XhoI; cDNA made by oligo-dT priming;
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Library constructed by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."

BASE COUNT      236 a    235 c    244 g    222 t      5 others
ORIGIN

Query Match           78.0%;    Score 895.2;   DB 13;   Length 942;
Best Local Similarity 97.4%;    Pred. No. 2.7e-34;
Matches 917;  Conservative    0;  Mismatches  23;   Indels   1;   Gaps  1;

Qy    182 GCCTATAGACAGCTGGCAGTGATGGTTCATCCTGACAAAATCATCATCCCGGGCTGAG 241
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db      1 GCCTATAGACAGCTGGCAGTGATGGTTCATCCTGACAAAATCATCATCCCGGGCTGAG 60

Qy    242 GAGGCCTTCAAGGTTTGCGAGCAGCTTGGGACAGCTTGCAGCAATGCTGAAAAGCGAAAG 301
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GAGGCCTTCAAGGTTTGCGAGCAGCTTGGGACAGCTTGCAGCAATGCTGAAAAGCGAAAG 120

Qy    302 GAGTATGAGATGAAAACGAATGGCAGAGAATGAGCTGAGCCGGTCAGTAAATGAGTTTCTG 361
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GAGTATGAGATGAAAACGAATGGCAGAGAATGAGCTGAGCCGGTCAGTAAATGAGTTTCTG 180

Qy    362 TCCAAGCTGCAAGATGACCTCAAGGAGGAGCAATGAATACTGAATGATGTGTAGCCGATGCCAA 421
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TCCAAGCTGCAAGATGACCTCAAGGAGGAGCAATGAATACTGAATGATGTGTAGCCGATGCCAA 240

Qy    422 GGAAAGCATAGGAGGTTTGAAATGGACCGGGAACCTAAGAGTGCCAGATACTGTGCTGAG 481
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GGAAAGCATAGGAGGTTTGAAATGGACCGGGAACCTAAGAGTGCCAGATACTGTGCTGAG 300

Qy    482 TGTAATAGGCTGCATCCGTCTGCTGGAGGAGACTTTGGGCAGAGTCAAGCATGTTGGGC 541
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TGTAATAGGCTGCATCCGTCTGCTGGAGGAGACTTTGGGCAGAGTCAAGCATGTTGGGC 360

Qy    542 CTCAAGATCACCTACTTTGCACTGATGGATGGAAGGTGTATGACATCACAGAGTGGGCT 601
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CTCAAGATCACCTACTTTGCACTGATGGATGGAAGGTGTATGACATCACAGAGTGGGCT 420

Qy    602 GGATGCCAGCGTGTAGGTATCTCCCAGTATCCCCAGATACCCAGAGTCCCCTATCACATCTCATTT 661
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GGATGCCAGCGTGTAGGTATCTCCCAGTATCCCCAGATACCCAGAGTCCCCTATCACATCTCATTT 480

Qy    662 GGTTCTCGGATTCCAGGCACCACCAGAGGGGCGGCAGAGAGCGCAGGGCAGAGCGCCCTCCTGCT 721
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GGTTCTCGGATTCCAGGCACCACCAGAGGGGCGGCAGAGAGCGCAGGGCAGAGCGCCCTCCTGCT 540

Qy    722 GATCTTCAGGATTTCTTGAGTTCGGATCTTTCAAGTACCCCAGGGCAGATGCCCAATGGG 781
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 GATCTTCAGGATTTCTTGAGTTCGGATCTTTCAAGTACCCCAGGGCAGATGCCCAATGGG 600

Qy    782 AACTTCTTGCAGCTCCTCAGCCTGCCCTGCCCTGGACGCCTCTAAGCCCAACAGC 841
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 AACTTCTTGCAGCTCCTCAGCCTGCCCTGCCCTGGACGCCTCTAAGCCCACAGC 660

Qy    842 ACAGTACCCAAGGGAGAAGCCAAAACCTAAGCGGCGGAGGGGAAGAAAGTGAGGAGGGCCCTTCAA 901
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 ACAGTACCCAAGGGAGAAGCCAAAACCTAAGCGGCGGAGGGGAAGAAAGTGAGGAGGGCCCCTTCAA 720

Qy    902 CGTTGATGCCCCTTCTCTTTCCTCAAATCAATGTCAGGGAGTCAAAAGGGCTGTAGCACA 961
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 CGTTGATGCCCCTTCTCTTTCCTCAAATCAATGTCAGGGAGTCAAAAGGGCTGTAGCACA 780

Qy    962 GGATGGGAGTTGATTTATCCCTCCTCCCCCAACACCTAGGAACTGAATCTTTT-CTTTT 1020
              |||||||||||||||||||||||||||||||||||||||||| ||||||||| |||||
Db    781 GGATGGGAGTTGATTTATCCCTCCTCCCCCAANANCTAAGAACTGAAACTTTTTCCTTTT 840

Qy   1021 TATTTTTGAGATGGAGTCTTGCTGTGTCCCAGCTGCAGTGCAGTGTGATCTCAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 TATTTTTGAGATGGAGTCTTGCTGTGTGTCCCANCTGCAGTGCAGTGTGATCTTCA 900

Qy   1081 CTTACTGCAACCTCTGTCTCCCGGGTTCAAGCAATTCCCC 1121
              |||||||||||||||||||||||||||||| |||||||||
Db    901 GCTACTGCCACCTCTGTCTCCCGGGTTCCCAGCAATTCTCCC 941


RESULT 2
AK017830
LOCUS       AK017830              2300 bp    mRNA    linear    HTC 05-DEC-2002
DEFINITION  Mus musculus 8 days embryo whole body cDNA, RIKEN full-length
            enriched library, clone:5730551F12 product:DOPAMINE RECEPTOR
            INTERACTING PROTEIN homolog [Rattus norvegicus], full insert
            sequence.
ACCESSION   AK017830
VERSION     AK017830.1  GI:12857283
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,K., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M. Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system-384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
            Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
            Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
            Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
            Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
            Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
            Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
            Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
            Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
            Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
            Carninci,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
            Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
            Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
            Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
```

REFERENCE
AUTHORS   Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
          Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
          Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
          and Hayashizaki,Y.
TITLE   Functional annotation of a full-length mouse cDNA collection
JOURNAL   Nature 409 (6821), 685-690 (2001)
MEDLINE   21085660
PUBMED   11217851
REFERENCE 5
AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
          Group Phase I & II Team.
TITLE   Analysis of the mouse transcriptome based on functional annotation
          of 60,770 full-length cDNAs
JOURNAL   Nature 420, 563-573 (2002)
REFERENCE 6 (bases 1 to 2300)
AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Arai,A., Aono,H.,
          Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Fukunishi,Y.,
          Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K.,
          Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M.,
          Kasukawa,T., Hori,H., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M.,
          Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K.,
          Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C.,
          Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
          Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y.,
          Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T.,
          Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K.,
          Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE   Direct Submission
JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
          Physical and Chemical Research (RIKEN), Laboratory for Genome
          Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
          RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
          Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
          URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
          Fax:81-45-503-9216)
COMMENT   Please visit our web site (http://genome.gsc.riken.go.jp/) for
          further details.
          cDNA library was prepared and sequenced in Mouse Genome
          Encyclopedia Project of Genome Exploration Research Group in Riken
          Genomic Sciences Center and Genome Science Laboratory in RIKEN.
          Division of Experimental Animal Research in Riken contributed to
          prepare mouse tissues. First strand cDNA was primed with a primer
          [5' GAGAGAGAGAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTTTTTVN 3'], cDNA was
          prepared by using trehalose thermo-activated reverse transcriptase
          and subsequently enriched for full-length by cap-trapper. cDNA went
          through one round of subtraction to Rot = 100.0. Second strand cDNA
          was prepared with the primer adapter of sequence [5'
          GAGAGAGAGAGGATTCGAGTAAATTAATTAATCCCCCCCCCCC 3']. cDNA was cleaved
          with BamHI and XhoI. Vector: a modified pBluescript KS(+) after
          bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3'
          end: BamHI. Host: DH10B.
FEATURES   Location/Qualifiers
     source   1..2300
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:5730551F12"
                     /db_xref="MGI:1907484"
                     /db_xref="taxon:10090"
                     /clone="5730551F12"
                     /tissue_type="whole body"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="8 days embryo"
     CDS   97..2208
                     /note="unnamed protein product; DOPAMINE RECEPTOR
                     INTERACTING PROTEIN homolog [Rattus norvegicus]
                     (SPTR|Q925G7, evidence: FASTY, 92.7%ID, 100%length,
                     match=2109)
                     putative"
                     /codon_start=1
                     /protein_id="BAB30962.1"
                     /db_xref="GI:12852284"
                     /db_xref="MGI:1921580"

          /translation="MAQKHPGERRLCGAHRSGGTSLSLSTSGSSVDPEILSFSGLRDSAE
          TAPNGTRCLKEHSGPKYTQPPNPAHWSDPSHGPPRGPGPPRGGGYPDESETGSEESGV
          DQELSRENETGYQEDGSPSFLSIPSACNCGSGPSGTYSEEGDGSSSSLCHHCTSP
          ALGEDEELEEEYDDEEPLKFPSDFSRVSSGKKFLSRROKHRFLKEDVRDSGRREPKA
          PGHRLARKRSQTDKRRGLGLWGVEELCQLGQAGFWWLIEBLLVLVGEYVETCGHLIYA
          CRKLKGSDLDLFRVWVGVWARRLGGWARMFQLSQSFFCVVGLLIRIRLVVGAFLL
          ALALFLGCLQLGKRFSVGLGNRLGWRDKTAWLFSWLGSPALHHCLTLLKDSRPWQQLV
          RLIQWGWQELPWVKQRTKKQNAPVASGRYCQPEEVTRLLTMAGVPEDELNPFHVLG
          VEATASDTELKKAYRQLAVMVHPDKNHHPRAEEAFKILRAAWDIVSNPERREVEMKR
          MAENELSRSVNEFLSKLQDDLKEAMNTMCSRCQGKHRRFEMDREPKSARYCAECNRL
          HPAEEGDFWAESSMLGLKITYFALMDGKVYDITEWAGCQRVGISPDTHRVPYHISFGS
          RVPGTSGRQRATPESPPADLQFLSRIFQVPPGMSNGNFFAAHPGPGTTSRPNS
          SVPKGEAKPKRKKVRPFQR"

BASE COUNT   557 a   560 c   675 g   508 t
ORIGIN

Query Match   69.5%;   Score 796.8;   DB 11; Length 2300;
Best Local Similarity 89.5%;   Pred. No. 5.8e-30;
Matches 882; Conservative   0; Mismatches 97; Indels   7; Gaps   2;

Qy   1 GCCTTGGGTCAAGCAGAATATTAATAGGCAGGGAATGCACCTGTAGCTAGTGGGGCTA 60
           ||||||||||||||||||||| || |||||| |||||||||||||||||||||||
Db   1302 GCCCTGGGTCAAACAGAGGACTAAGAAACAGGGCAATGCACCTGTAGCTAGCGGGGCATA 1361

Qy   61 CTGCCAGCCTGAAGAGGAAGTGGCTCGACTCTTGACCATGCTGCGGGTTCCTGAGGATGA 120
           |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db   1362 CTGCCAGCCTGAAGAGGAAGTGAACCCGACTCTTGACCATGCTGCGGGTTCCTGAGGATGA 1421

Qy   121 GCTAAACCCTTCCATGTACTGGGGTTGAGCGCCACAGCATCAGATGTTGAACTGAAGAA 180
           |||| |||| ||| |||||| ||||| |||||||| |||||||| | | |
Db   1422 ACTAAACCCTTTCATGTGCTGGGGTTGAAGCTACACATCCGACACTGAACTAAAGAA 1481

Qy   181 GGCCTATAGACAGTCGGCAGTGATGTTCATCCTGACAAAATCATCCCGGGCTGA 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1482 GGCCTATAGGCAGTCAGCAGTAATGTCCATCCTGATAAAAAATCACCATCCCGGGCTGA 1541

Qy   241 GGAGGCCTTCAAGGTTTCGGAGCAGCTTGGGACATTGTCGCAATGCTGAAAAGCGAAA 300
           |||||||||||||||||| ||||||||| ||||||||| ||| ||||||| |
Db   1542 GGAGGCCTTCAAAATTTCGGGGCAGCTTGGGACATTGTCAGCAACCCAGAGAGGCGAA 1601

Qy   301 GGAGTATGAGATGAAACGAATGCAGAGAATGAGCTGGAGCCGGTCAGTAAGTGAGTTTCT 360
           |||| |||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db   1602 GGAATATGAGATGAAACGAATGCAGAGAATGAGCTGAGCCGGTCAGTCAGTGAATGAGTTTCT 1661

Qy   361 GTCCAAGCTGCAAGATGACCTCAAGGAGGCAATGAATACTATGATGTGTAGCCGATGCCA 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1662 GTCGAAACTACAAGATGACCTCAAGGAGGCAATGAACACGATGATGTGCAGCAGATGCCA 1721

Qy   421 AGGAAAGCATAGGAGGTTTGAAATGGACCCGGAACCTAAGAGTGCCAGATACTGTGCTGA 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1722 AGGAAAGCATAGGAGGTTTGAAATGGACCCGGAACCTAAGAGTGCCAGATACTGTGCTGA 1781

Qy   481 GTGTAATAGGCTGCATCCTGCTGAGGAAGGAGACTTTGGGCAGAGTCAAGCATGTTGGG 540
           |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1782 GTGTAACAGGCTGCATCCTGCTGAGGAAGGAGACTTTGGGCAGAGTCGCAGCATGCTGGG 1841

Qy   541 CCCAAGATCACCTACTTTGCACTGATGGATGGAAAGGTGTATGACATCACAGAGTGGGGC 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1842 CCTCAAGATCACTTACTTTGCGCTGATGGATGGAAAGGTGTATGACATCACAGAGTGGGGC 1901

Qy   601 TGGATGCCAGCGTGTAGGTATCTCCCCAGATACCCACAGAGTCCCTATCACATCCATT 660
           ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1902 TGGATGCCAGCGTGTAGGTATCTCCCCAGATACTCACAGAGTTCCTTACCACATCCATT 1961

Qy   661 TGGTTCTCGGATTCCAGGCACCACGTCCGGACCCAGGGCGGCACCCCAGATGCCCCTCCGC 720
           |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db   1962 TGGTTCTCGGGATTCCCGGTACCCGGACCAGTGGGCCCACTCCAGAGTTCCCCTCCTGC 2021

Qy   721 TGATCTTCAGGATTTCTTGAGTCCGGATCTTCAAGTACCCCAGGGCAGAGATGCCCAATGG 780
           ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2022 TGACCTGCAGGATTTCTTGAGCCGGATCTTCAAGTACCCCGGGGCCGGATGTCCAATGG 2081

Qy   781 GAACTTCTTTGCAGCTCTCCAGCCTGCCCCTGACCCCTGAGCCCCTGCAGCCTCTTAAGCCCAACAG 840

```
Db    2082  GAACTTCTTGCGCACCTCACCCTGGCCCTGGGACCCACTTCGACCTCTAGGCCCAACAG 2141
            |||||||||  ||||||  ||| |||||||||||||| |||||||||||  |||||||||
Qy     841  CACAGTACCCAAGGGAGAAGCCAAACCTAAGCGGCGGAGAAAGTGAGGAGGCCCTTCCA  900
            |||||||||  ||||||  ||| |||||||||||||| |||||||||||  |||  |||
Db    2142  TTCAGTACCCAAGGGAGAAGCCAAACCTAAACGGCGGAGAAAGTGAGGCGGCCCTTTCA 2201

Qy     901  ACGTTGATGCCCTTCT---TTCCTCAAATCAATGTCAGGGAGTCAAAAGGGCTGT-- 955
            |||||||  ||||||||   |||||||||||||||||||||||||||||||||||
Db    2202  ACGATGACACCCCTTCTCTTCCTCCTCAAATCAATGTCAGGGAGTCAAAAGGGCTGTGT 2261

Qy     956  --AGCACCAGGATGGAGTTTGATTTAT  979
              |||||||||||||||| ||||||||
Db    2262  ACAGGCACCAGGATGGAATTTGATTTAT 2287
```

```
RESULT 3
AK045445
LOCUS       AK045445                 2712 bp    mRNA    linear    HTC 05-DEC-2002
DEFINITION  Mus musculus adult male corpora quadrigemina cDNA, RIKEN
            full-length enriched library, clone:B230120I06 product:DOPAMINE
            RECEPTOR INTERACTING PROTEIN homolog (Rattus norvegicus), full
            insert sequence.
ACCESSION   AK045445
VERSION     AK045445.1  GI:26337372
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,K., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system -384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
            Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
            Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
            Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
            Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
            Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
            Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
            Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
            Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
            Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
            Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
            Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
            Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
            Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
            Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
```

```
            Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
            Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
            Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
            and Hayashizaki,Y.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409 (6821), 685-690 (2001)
  MEDLINE   21085660
   PUBMED   11217851
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 2712)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JUL-2001) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Insitute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohana,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.go.jp/
            URL:http://fantom.gsc.riken.go.jp/.
FEATURES             Location/Qualifiers
     source          1..2712
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:B230120I06"
                     /db_xref="taxon:10090"
                     /clone="B230120I06"
                     /sex="male"
                     /tissue_type="corpora quadrigemina"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     CDS             177..2288
                     /note="unnamed protein product; DOPAMINE RECEPTOR
                     INTERACTING PROTEIN homolog [Rattus norvegicus]
                     (SPTR|Q925G7; evidence: FASTY, 92.7%ID, 100%length,
                     match=2109]
                     /codon_start=1
                     /protein_id="BAC32272.1"
                     /db_xref="GI:26337373"
                     /translation="MAQKHPGERRLCGAHRSGGTSLSTGGSVDPEILSFSGLRDSAE
                     TAPNGTRCLKEHSGPKYTQPPNPAHWSDPSHGPPGRPGPRGGGYPDESETGSEESGV
                     DQELSRENETGYOEDGSPSFLSIPSACNCQGSPGVPEGTYSEEGDGSSSLCHHCTSP
                     ALGEDEELEEYDDEEPLKFPSDFSRVSSGKKPLSRRQKHRFLIKEDVRDSGRREPKA
                     PGRHRLARKRSQTDKRRGLGLVGEELCQLGQAGFWLIELLVLVGEYVETCGHLIYA
                     CRKLKGSDLDLFRVWVGWARRLGGWARMFQFLSQSFFCVVGLLIRLRVVGAFLLL
                     ALALFLGCLQLGWRFSVGLGNRLGWRDKTAWLFSWLGSPALHHCLTLLKDSRPWQQLV
                     RLIQWGWQEPLPWVKQRTKKQGNAPVASGRYCQPEEVTLLTMAGVPEDELNPFFVLG
                     VEATASDTELKKAYRQLAVMVHPDKNHPRAEEAFKILRAAWDIVSNPERRKEYEMKR
```

```
Db   2282  ACGATGACCACCCCTTCTTCCTTCCTCCTCAAATCAATGTCAGGGAGTCAAAAGGGCTGTGT 2341
                                                              |||||||||||||||||||
Qy    956  --AGCCACAGGATGAGTTTGATTAT 979
             |||||||||||||||
Db   2342  ACAGCCACAGGATGGAATTTGATTAT 2367
```

RESULT 4
AK032681

```
LOCUS       AK032681                2639 bp    mRNA    linear   HTC 05-DEC-2002
DEFINITION  Mus musculus 10 days neonate cerebellum cDNA, RIKEN full-length
            enriched library, clone:6530416L18 product:DOPAMINE RECEPTOR
            INTERACTING PROTEIN homolog [Rattus norvegicus], full insert
            sequence.
ACCESSION   AK032681
VERSION     AK032681.1  GI:26082970
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530013
   PUBMED   11076861
REFERENCE   4
  AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
            Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
            Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
            Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
            Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
            Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
            Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
            Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
            Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
            Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
            Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
            Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
            Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
            Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
            Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
            Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
            Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
            Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
            and Hayashizaki,Y.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409 (6821), 685-690 (2001)
  MEDLINE   21085660
   PUBMED   11217851
REFERENCE   5
```

```
MAENELSRSVNEFLSKLQDDLKEAMNTMWCSRCQKHRFEMDREPKSARYCAECNRL
HPAEEGDFWAESSMLGLKITYFALMDGKVDITEWAGCQRVGISPDTHRVPYHiSFGS
RVPGTSGRGRATPESPPADLQDFLSRIFQVPPGPMSNGNFFAAPHPGGTTSTSRPNS
SVPKGEAKPKRKKVRRPFQR"
```

```
BASE COUNT   664 a   663 c   778 g   607 t
ORIGIN

Query Match          69.5%;  Score 796.8;  DB 11;  Length 2712;
Best Local Similarity 89.5%;  Pred. No. 5.2e-30;
Matches  882;  Conservative  0;  Mismatches  97;  Indels  7;  Gaps  2;

Qy     1  GCCTTGGTCAAGCAGAGATATTAATAGGCAGGGAATGCACCTGTAGCTAGTGGCGCCTA    60
          ||||| ||||||| |||||  | | || ||| ||||||||||||||||||| ||  | ||
Db  1382  GCCCTGGGTCAAACAGAGGACTAAGAAACAGGGCAATGCACCTGTAGCTAGCGGGGCGATA 1441

Qy    61  CTGCCAGCCTGAAGAGGAAGTGGCTCGACTCTTGACCATGGCTGGGGTTCCTGAGGATGA  120
          ||||||||||||||||||||||  | |||||||||||||||| |||||||||||| |||||
Db  1442  CTGCCAGCCTGAAGAGGAAGTGACCCGACTCTTGACCATGGGTGGGGTTCCTGAAGATGA 1501

Qy   121  GCTAAACCCTTCCATGTACTGGGGGTTGAGCGCCAGCATCAGATGTTGAACTGAAGAA   180
          ||||||||||| ||||| |||||||||||   |||||||| |||||| || |||||||||
Db  1502  ACTAAACCCTTTCCATGTGCTGGGGGTTGAAGCTACAGCATCCGACACTGAACTAAAGAA 1561

Qy   181  GGCCTATAGACAGCTGCGAGTGATGGTTCATCCTGACAAAAATCATCATCCCGGCTGA   240
          |||||||||||||| |||| ||||||| ||||||| ||||||||| |||||| ||||||
Db  1562  GGCCTATAGGCAGCGCAGCAGTAATGGTCCATCCTGATAAAAATCACCATCCCCGGGCTGA 1621

Qy   241  GGAGGGCCTTCAAGGTTTGCGAGCAGCTTGGGACATTGTCAGCAATGCTGAAAAGCGAAA  300
          ||||||||||||| |||||||  |||||||||||||| |||||||||||||||| ||| ||
Db  1622  GGAGGGCCTTCAAAATTTGCGGGCAGCTTGGGACATTGTCAGCAATGCTGAGAGGCCGAA 1681

Qy   301  GGAGTATGAGATGAAACGAATGCGAGAGAATGAGCTGAGCCGGTCAGTAAATGAGTTTCT  360
          ||| |||||||||||||| ||||||||||||||||| |||||||||||| |||||||| |
Db  1682  GGAATATGAGATGAAACGGATGCGAGAGAATGAGCTCAGCCGGTCAGTGAATGAGTTTCT 1741

Qy   361  GTCCCAAGCTGCGAAGATGACCTCAAGGAGGCAATGAATAACTATGATGTGTAGCCGATGCCA  420
          ||| | |||| | ||||||||||||||||||||||||| |||  | |||| || ||||| |||
Db  1742  GTCGAAACTACAAGATGACCTCAAGGAGGCAATGAACACGATGAGTGCGCAGAGTGCCA 1801

Qy   421  AGGAAAGCATAGGAGGTTGAAATGGACCGGGAACCTAAGAGTGCCAGATACTGTGTGA   480
          |||||||||||||||||||||||||||| |||||| |||||||||||||||||||||||
Db  1802  AGGAAAGCATAGGAGGTTGAAATGGACCGGGAACCCAAGAGTGCCAGATACTGTGTGA 1861

Qy   481  GTGTAATAGGCTGCATCCTGCTGAGGAAGGAGACTTTTGGGCAGAGTCAAGCATGTTGGG  540
          ||||  ||||||| ||| ||||| ||||||||||||||||| || || ||||||| ||||
Db  1862  GTGTAACAGGCTGCATCCTGCTGAGGAAGGAGACTTTTGGGCAGAGTCGAGCATGCGGG 1921

Qy   541  CCTCCAAGATCACCTACTTTGCACTGATGGAAAGGTGTATGAACATCACAGAGTGGGC   600
          |||||||||||||||||||||| |||||||||||||| |||||| |||||||||||| |
Db  1922  CCTCCAAGATCACCTTACTTTGCGCTGATGGAAAGGTGTATGACATCACAGAGTGGGC 1981

Qy   601  TGGGATGCCAGCGTGTAGGTATCTCCCCAGATACCCACCAGAGCCCCTATCACATCCATT  660
          |||||||||||||||||| |||||||| |||| ||||||||||||| ||  |||| ||||
Db  1982  TGGATGCCAGCGTGTGGTATCTCCCCAGATACTCACCAGAGCCCCTTACCACCATCCATT 2041

Qy   661  TGGTTCTCGGATTCCAGGCACCAGAGGGGCGCAGAGCCCACCCCAGATGCCCCTCCTGC  720
          |||||||||| ||||||| |||||| ||||| ||||| ||||||||||||||  |||||
Db  2042  TGGTTCTCGGGTACCCGGCACCAGTGGTGCCGCCGGAGTCCCACTCCCAGAGTCCCTCCTGC 2101

Qy   721  TGATCTTCAGGATTCTTGAGTCGGATCTTTCAAGTACCCCCCAGGGCAGATGGCCAATGG  780
          |||| ||||| ||||||||||| |||| |||||| |||| || |||||||  ||||||||
Db  2102  TGACCTGCAGATTCTTGAGCCGGGATCTTTCAAGTACCTCCGGGGCCGATGTCCAATGG 2161

Qy   781  GAACTTCTTTGCAGCCTCCTCAGCCGGTTGCAGCCTCAAGCCCAACAG   840
          ||||||||||| || |||||| |||||| ||||||| ||| |||||||
Db  2162  GAACTTCTTTGCCGCACCTCACCCTGCCACCACTTCGACCTCTAGGCCCAACAG 2221

Qy   841  CACAGTACCCCAGGGAGAAGCAAACCTAAGCGCGGAAGCCGGAAGAAAGTGAGAGAACCTTCCA  900
          ||||| |||||||||||||| |||||| |||||||||||||| ||||||||||||| ||||||
Db  2222  TTCAGTACCCCAAGGAGAAGCAAACCTAAACGGGCCGAAGAAAGTGAGGCGCGCAAAAGGGCTGT-- 2281

Qy   901  ACGTTGATGCCCCTTCTCT---TTCCTCAAATCAATGTCAGGGAGTCAAAAGGGCTGT--  955
          ||| || |||  |||||
```

AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
          Group Phase I & II Team.
TITLE'   Analysis of the mouse transcriptome based on functional annotation
          of 60,770 full-length cDNAs
JOURNAL   Nature 420, 563-573 (2002)
REFERENCE 6 (bases 1 to 2639)
AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
          Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
          Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
          Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
          Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
          Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
          Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
          Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
          Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
          Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
          Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
          Muramatsu,M. and Hayashizaki,Y.
TITLE   Direct Submission
JOURNAL   Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
          Physical and Chemical Research (RIKEN), Laboratory for Genome
          Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
          RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
          Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
          URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
          Fax:81-45-503-9216)
COMMENT   cDNA library was prepared and sequenced in Mouse Genome
          Encyclopedia Project of Genome Exploration Research Group in Riken
          Genomic Sciences Center and Genome Science Laboratory in RIKEN.
          Division of Experimental Animal Research in Riken contributed to
          prepare mouse tissues.
          Please visit our web site for further details.
          URL:http://genome.gsc.riken.go.jp/
          URL:http://fantom.gsc.riken.go.jp/.
FEATURES   Location/Qualifiers
     source   1..2639
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:6530416L18"
                     /db_xref="taxon:10090"
                     /clone="6530416L18"
                     /tissue_type="cerebellum"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="10 days neonate"
     misc_feature   64..2176
                     /note="DOPAMINE RECEPTOR INTERACTING PROTEIN homolog
                     [Rattus norvegicus] (SPTR|Q92SG7, evidence: FASTY,
                     92.7%ID, 100%length, match=2109)
                     putative"
BASE COUNT   666 a   634 c   738 g   601 t
ORIGIN

Query Match   69.3%; Score 795.2; DB 11; Length 2639;
Best Local Similarity 89.4%; Pred. No. 6.3e-30;
Matches 881; Conservative 0; Mismatches 98; Indels 7; Gaps 2;

Qy   1 GCCTTGGGTCAAGCAGATATTAATAGGCAGGGAATGCACCTGTAGTGGCCCTA 60
          ||||||||||||||||||||| ||| ||||||||||||||||| ||||||||||
Db   1270 GCCCTGGGTCAAACAGGACTAAGAAACAGGGCAATGCACCTGTAGTAGCGGGCATA 1329

Qy   61 CTGCCAGCCTGAAGAGGAAGTGGCTCGACTCTTGACCATGGCTGGGGTTCCTGAGGATGA 120
          ||||||||||||||||||||| ||||||||||| |||||||||||||||||||||
Db   1330 CTGCCAGCCTGAAGAGGAGAGTGACCCGACTCTTGACCATGGCTGGGGTTCCTGAAGATGA 1389

Qy   121 GCTAAACCCTTTCCATGTACTGGGGGTTGAGGCCACCAGCATCAGATGTTGACTGAAGAA 180
          ||||||||||||| ||||| |||||||||||||||| ||||||| ||||||| ||||||
Db   1390 ACTAAACCCTTTCATGTGCTGGGGGTTGAAGCTACAGCATCCGACACTGAACTAAAGAA 1449

Qy   181 GGCCTATAGACAGCTGGCAGTGATGGTTCATCCTGACAAAAATCATCATCCCCGGGCTGA 240
          |||||||||||||||||||||| |||||||| |||||||||||||| ||||||||||||
Db   1450 GGCCTATAGGCAGCTAGCAGCAGTAATGGTTCCATCCTGATAAAAATCACCATCCCCGGGCTGA 1509

Qy   241 GGAGGCCTTCAAGTTTTGCGAGCAGCTTGGGACAATTGTCAGCAATGCTGAAAAAGCGAAA 300
          |||||||||||||||||||||||||||||||||||||||||||||||| |||| |||
Db   1510 GGAGGCCTTCAAAATTTTGCGGGCAGCTTGGGACAATTGTCAGCAACCCAGAGAGGGGAA 1569

Qy   301 GGAGTATGAGATGAAACGAATGGCAGAGAATGAGCTGAGCCGGTCAGTAAATGAGTTTCT 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1570 GGAATATGAGATGAAACGGATGGCAGAGAATGAGCTCAGCCGGTCAGTGAATGAGTTTCT 1629

Qy   361 GTCCAAGTGCAAGATGACCTCAAGGAGCAATGAATACTATGATGTGTAGCCGATGCCA 420
          ||| |||||||||| ||||||||||||||||||||||| |||||||||||||||||||
Db   1630 GTCGAAACTACAAGATGACCTCAAGGAGGCAATGAACACGATGATGTGCAGCAGATGCCA 1689

Qy   421 AGGAAAGCATAGGAGGTTTGAAATGACCGGAACCTAAGAGTGCCAGATACTGTGCTGA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1690 AGGAAAGCATAGGAGGTTTGAAATGACCGGGAACCCAAGAGTGCCAGATACTGTGCTGA 1749

Qy   481 GTGTAATAGGCTGCATCCTGCTGAGGAAGGAGACTTTTGGCAGAGTCAAGCATGTTGGG 540
          ||||||| |||||||||||||||||||||||||| |||||||||||||||||||||||
Db   1750 GTGTAACAGGCTGCATCCTGCTGAGGAAGGAGAGACTTTTGGGCAGAGTCGAGCATGCTGGG 1809

Qy   541 CCTCAAGATCACCTACTTTGCACTTGCACATGATGAGAAGGTGTATGACATCACAGAGTGGGC 600
          ||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db   1810 CCTCAAGATCACTTACTTTGCGCTGATGATGAGAAGGTGTATGACATCACAGAGTGGGC 1869

Qy   601 TGGATGCCAGCGTGTAGGTATCTCCCAGATACCCACAGAGTCCCTATCACATCTCATT 660
          |||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db   1870 TGGATGCCAGCGTGTGGGTATCTCCCAGATACTCACAGAGTTCCTTACCACATCTCATT 1929

Qy   661 TGGTTCTCGGATTCCAGGCACCAGGGGCGGCAGAGCCACCCCAGATGCCCCTCCTGC 720
          |||||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db   1930 TGGTTCTCGGGTACCCGGCCACCAGTTGGCCCGGCCGACCCGCCCCAGTTCCCAGAGTCCCCTGC 1989

Qy   721 TGATCTTCAGGATTCTTGAGTCGGATCTTCAAGTACCCCAGGGCAGATGCCAATGG 780
          |||||||||||||||||||||||| |||||||||||||||||||||| ||||||
Db   1990 TGACCTGCAGGATTTCTTGAGCCCGGATCTTCAAGTACCTCCGGGGCCGATGTCCAATGG 2049

Qy   781 GAACTTCTTTGCAGTCCTCAGCCGCGCCCCGAGCGTCGCAGCCTTCTAAGCCACACAG 840
          |||||||||| |||||||||||||||| |||||||||||||||||||||||||||||
Db   2050 GAACTTCTTTGCCGCACCTCACCCTGGCCTGGGCCTCAGCCTCTAGGCCCCAACAG 2109

Qy   841 CACAGTACCCAAGGAGAAGCCAAACCTAAGCGGCGGGAAGAAGTGAGAGGCCCTTCCA 900
          ||||||||||||||||||||| ||||||||||||||||||| |||||||||||||
Db   2110 TTCAGTACCCAAGGGAGAAGCCAAACCTAAAGGGCGGGAAGAAGTGAGGGGCGCCCCTTTCA 2169

Qy   901 ACGTTGATGCCCCTTCTCT---TTCCTCAAATCAATGTCAGGGAG!CAAAAGGGCTGT-- 955
          |||||| |||||||||| ||||||||||||||||||||||||||| ||||||||||||
Db   2170 ACGATGACACCCCTTTCTCCTTCCTCAAATCAATCAATGTCAGGGAGTCAAAAGGGCTGTGT 2229

Qy   956 --AGCACAGGATGGAGTTGATTTAT 979
          ||||||||||||||||||||||||
Db   2230 ACAGCACAGGATGGAGTTGATTTAT 2255

RESULT 5
AK082579
LOCUS   AK082579   4231 bp   mRNA   linear   HTC 05-DEC-2002
DEFINITION Mus musculus 0 day neonate cerebellum cDNA, RIKEN full-length
            enriched library, clone:C230066C20 product:DOPAMINE RECEPTOR
            INTERACTING PROTEIN homolog [Rattus norvegicus], full insert
            sequence.
ACCESSION   AK082579
VERSION   AK082579.1 GI:26349792
KEYWORDS   HTC; CAP trapper.
SOURCE   Mus musculus (house mouse)
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE   High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
  PUBMED   10349636

REFERENCE   2
AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE   Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE   20499374
PUBMED   11042159
REFERENCE   3
AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Wataihiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE   RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE   20530913
PUBMED   11076861
REFERENCE   4
AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
            Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
            Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
            Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
            Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
            Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
            Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
            Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
            Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
            Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
            Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
            Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
            Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
            Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
            Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
            Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
            Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
            Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
            and Hayashizaki,Y.
TITLE   Functional annotation of a full-length mouse cDNA collection
JOURNAL   Nature 409 (6821), 685-690 (2001)
MEDLINE   21085660
PUBMED   11217851
REFERENCE   5
AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
TITLE   Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6 (bases 1 to 4231)
AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,J., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
TITLE   Direct Submission
JOURNAL   Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT   cDNA library was prepared and sequenced in Mouse Genome

            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.go.jp/.
            URL:http://fantom.gsc.riken.go.jp/.
FEATURES   Location/Qualifiers
     source   1..4231
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:C230066C20"
                     /db_xref="taxon:I0090"
                     /clone="C230066C20"
                     /tissue_type="cerebellum"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="0 day neonate"
     CDS   84..2195
                     /note="unnamed protein product; DOPAMINE RECEPTOR
                     INTERACTING PROTEIN homolog [Rattus norvegicus]
                     (SPTR|Q925G7, evidence: FASTY, 92.7%ID, 100%length,
                     match=2109)
                     putative"
                     /codon_start=1
                     /protein_id="BAC38536.1"
                     /db_xref="GI:26349793"
                     /translation="MAQKHPGERRLCGAHRSGGTSLSTSGSSVDPEILSFSGLRDSAE
                     TAPNGTRCLKEHSGPKYTQPPNPAHWSDPSHGPGPRGPGGYPDESTGSEESGV
                     DQELSRENETGYOEDGSPSFLSIPSACNCQSPGVPEGTYSEEDGSSSSLCHHCTSP
                     ALGEDELEEEYDDEEPLKFPSDFSRVSSGKKPLSRRQKRHFLIKEDVRDSGRREPKA
                     PGRHRLARKRSQTDKRRGLGLWGVEELCQLQAGFWWLIELLVLWGEYVETCGHLIYA
                     CRKLKGSDLDLFRVWVGWARRLGGWARMMFQFLSQSFPCVGLLIRLRVGAFLL
                     ALALFLGCLQLGWRFSVGLGNRLGWRDXTAWLFSWLGSPALHHCLTLLKDSRPWQQLV
                     RLIQWGWQELPWVKQRTKKQNAPVASGRYCQPEEVTRLLTMAGVPEDLNPFHVLG
                     VRATASDTELKKAYRQLAVMVHPDKNHPRAEEAFKILRAAWDIVSNPERRKEYEMKR
                     MAENELSRVNEFLSKLQODLKEAMNTMMCSRCQGKHRFEMDREPKSAGYCAECNRL
                     HPAEEGDFWAESSMLGLKITYFALMDGKYDITEWAGCQRVGISPDTHRVPYHISFGS
                     RVPGTSGRQRATPESPPADLQDFLSRIFQVPPGPMSNGNFFAAPHPGGTTSTSRPNS
                     SVPKGEAKPKRKKVRRPFQR"
BASE COUNT   1092 a   948 c   1059 g   1132 t
ORIGIN

Query Match   69.3%; Score 795.2; DB 11; Length 4231;
Best Local Similarity   89.4%; Pred. No. 4.6e-30;
Matches 881; Conservative   0; Mismatches 98; Indels   7; Gaps   2;

Qy   1 GCCTTGGGTCAAGCAGAAGATATTAATAGGCAGGGGAATGCACCTGTAGCTAGTGGGCGCTA 60
            ||| |||||||||||||||||| |||| || |||||| |||||| ||||||||||| |||||
Db   1289 GCCCTGGGTCAAACAGAGGACTAAGAAACAGGGCAATGCACCTGTAGCTAGTGGGCGGCATA 1348

Qy   61 CTGCCAGCCTGAAGAGGAAGTGGCTCGACTCTTGACCATGGCTGGGGTTCCTGAGGATGA 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1349 CTGCCAGCCTGAAGAGGAAGTGAGACCCGACTCTTGACCATGGCTGGGGTTCCTGAAGATGA 1408

Qy   121 GCTAAACCCTTCCATGTACTACTGGGGTTGAGGCCACCAGCATCAGATGTTGAACTGAAGA 180
            |||||||||||||||||||||||||||||||||||||| ||||||||||||||||| |||||
Db   1409 ACTAAACCCTTTCATGTGCTGGGGTTGAAGCTACAGCATCCGACCACTGAACTAAAGAA 1468

Qy   181 GGGCTATAGACAGCTGGCAGTGATGGTTCATCCTGACAAAAATCATCATCCCCGGGCTGA 240
            ||||||||||||||||||||||||||||||||||||||||| |||||| |||||||||||||
Db   1469 GGGCTATAGGCAGCTAGCAGTAATGGTCCATCCTCGATAAAAATCACCATCCCCGGGCTGA 1528

Qy   241 GGAGGCCCTTCAAGGTTTTGCGAGCAGCTTGGGACATTGCTCAGCAATGCTCAACACCAGCG 300
            ||||||||| |||||| |||||||||||||||||||||||||||||||||||||||||||||
Db   1529 GGAGGCCTTCAAAAATTTGCGGGCAGCTTGGGACATTGTCGCAACCCAGAGAGGCGGGAA 1588

Qy   301 GGAGTGATGAGATGAAACGAATGGCAGAGAATGGCCGGTCAGTAAATGAGTTTCT 360
            |||||||||||||||||||| ||||| ||||| |||| |||||| |||||| |||||
Db   1589 GGAATATGAGATGAAACGATGGCAGAGAATGAGCTCAGCCGGTCAGTGAATGAGTTTCT 1648

Qy   361 GTCCAAGCTGCAAGATGACCTCAAGGAGGCAGAATGTGTAGCGATGCCATGCCA 420
            ||| ||| || ||||||| |||||||||||||||||| |||||||| || |||||

```
                                        /mol_type="mRNA"
                                        /db_xref="taxon:9606"
                                        /clone="IMAGE:6729715"
                                        /tissue_type="adenocarcinoma, cell line"
                                        /lab_host="DH10B (phage-resistant)"
                                        /clone_lib="NIH_MGC 107"
                                        /note="Organ: breast; Vector: pOTB7; Site_1: EcoRI;
                                        Site 2: XhoI; cDNA made by oligo-dT priming.
                                        Directionally cloned into EcoRI/XhoI sites using the
                                        following 5' adaptor: GGCACGAG(G). Library constructed by
                                        Ling Hong in the laboratory of Gerald M. Rubin (University
                                        of California, Berkeley) using ZAP-cDNA synthesis kit
                                        (stratagene) and Supersript II RT (Life Technologies).
                                        Note: this is a NIH_MGC Library."

BASE COUNT     234 a    228 c    238 g    218 t
ORIGIN

Query Match         69.1%;  Score 792.6;  DB 13;  Length 918;
Best Local Similarity 98.7%;  Pred. No. 1.7e-29;
Matches  820;  Conservative    0;  Mismatches    9;  Indels    2;  Gaps    2;
```

Right-hand alignment (Qy / Db):

```
Qy    234 GGGCTGAGGAGCCTCAAGGTTTCGAGCAGCTTGGACAGTTGTCAGCAATGCTGAAA    293
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GGGCTGAGGAGCCTCAAGGTTTCGAGCAGCTTGGACAGTTGTCAGCAATGCTGAAA     60

Qy    294 AGCGAAAGGAGTATGAGATGAAAACAATGACGCAGAGAATGAGCTGAGCCGGTCAGTAAATG 353
          |||||||||||||||||||||||||||||
Db     61 AGCGAAAGGAGTATGAGATGAAAACGATGAGGCAGATGAGCGGTGAGCCGGTCAGTAAATG  120

Qy    354 AGTTTCGTCCAAGTCGCAAGATGACCTCAAGGAGCAATGAATACTATGATGTGTAGCC    413
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AGTTTCGTCCAAGTCGCAAGATGACCTCAAGGAGCAATGAATACTATGATGTGTAGCC   180

Qy    414 GATGCCAAGGAAAGCATAGGAGGTTTGAAATGGACCGGGAACCTAAGAGTGCCAGATACT  473
          |||||||||||||||||||||||||||||||||
Db    181 GATGCCAAGGAAAGCATAGGAGGTTTGAAATGGACCGGGAACCTAAGAGTGCCAGATACT  240

Qy    474 GTGCTGAGTGTAATAGGCTGCATCCTCGCTGAGGAAGGAGACTTTGGGCAGAGTGATGACATCACAG 533
          ||||||||||||||||||||||||||||||
Db    241 GTGCTGAGTGTAATAGGCTGCATCCTCGCTGAGGAAGGAGACTTTGGGCAGAGTGATGACATCACAG 300

Qy    534 TGTTGGGCCTCAAGATCACCTTTGCACTGATGGATGGAAAGGTGTATGACATCACAG    593
          ||||||||||||||||||||||||||||||
Db    301 TGTTGGGCCTCAAGATCACCTTTGCACTGATGGATGGAAAGGTGTATGACATCACAG   360

Qy    594 AGTGGGCTGGATGCCAGCCGTGTAGGTATCTCCCCAGATACCCCACAGAGTCCCTATCCA  653
          ||||||||||||||||||||||||||||||
Db    361 AGTGGGCTGGATGCCAGCCGTGTAGGTATCTCCCCAGATACCCCACAGAGTCCCTATCCACA  420

Qy    654 TCTCATTTGGTTCTCGGATTCCAGGCACCAGAGAGCGGCGCACCCCAGATGCCC    713
          ||||||||||||||||||||||||||||||
Db    421 TCTCATTTGGTTCTCGGATTCCAGGCACCAGAGAGCGGCGCACCCCAGATGCCC   480

Qy    714 CTTCCTGCTGATCTTCAGGATTTCTTGAGTCGGATCTTTCAAGTACCCCCAGGGCAGATGC  773
          ||||||||||||||||||||||||||||||
Db    481 CTTCCTGCTGATCTTCAGGATTTCTTGAGTCGGATCTTTCAAGTACCCCCAGGGCAGAGATGC 540

Qy    774 CCAATGGAACTTCTTTGCAGCTCTGCAGCTGGAGCCGTGGAGCCGGTGCAGCCTCTAAGC  833
          ||||||||||||||||||||||||||||||
Db    541 CCAATGGAACTTCTTTGCAGCTCTGCAGCTGGAGCCGTGGAGCCGGTGCAGCCTCTAAGC  600

Qy    834 CCAACAGCACAGTACCCAAGGGAGAAGCCAAACCTAAGCGGCGGGAAGAAAGTGAGGAGGC 893
          ||||||||||||||||||||||||||||||
Db    601 CCAACAGCACAGTACCCAAGGGAGAAGCCAAACCTAAGCGGCGGGAAGAAAGTGAGGAGGC  660

Qy    894 CCTTCCAAGTTGATGCCCCTTCTGTTCTTCCTCAAATCAATGTCAGGGAGTCAAAAGGGCT 953
          ||||||||||||||||||||||||||||||
Db    661 CCTTCCAAGTTGATGCCCCTTCTGTTCTTCCTCAAATCAATGTCAGGGAGTCAAAAGGGCT  720

Qy    954 GTAGCACAGGATGGAGTTTGATTTATCCTCCCCAAACACCTAGGAACTGAATCTTT   1013
          ||||||||||||||||||||||||||||||
Db    721 GTAGCACAGGATGGAGTTTGATTTATCCTCCCCAAACACCTAGGAACTGAATCTTT    780
```

Left-hand alignment (Db / Qy):

```
Db   1649 GTCGAAACTACAAGATGACCTCAAGGAGCAATGAACACGATGATGTGCAGCAGATGCCA 1708
Qy    421 AGGAAAGCATAGGAGGTTTGAAATGGACCGGGAACCTAAGAGTGCCAGATACTGTGCTGA  480
Db   1709 AGGAAAGCATAGGAGGTTTGAAATGGACCGGGAACCCAGAGTGCCGGATACTGTGCTGA 1768
Qy    481 GTGTAATAGGCTGCATCCTCCTGAGGAAGGAGACTTTGGGCAGAGTCAAGCATGTTGGG  540
Db   1769 GTGTAACAGGCTGCATCCTCCTGCTGAGGAAGGAGACTTTGGGCAGAGTCGAGCATGCTGGG 1828
Qy    541 CCTCAAGATCACCTACTTTGCACTGATGGATGGAAAGGTGTATGACATCACAGAGTGGGC  600
Db   1829 CCTCAAGATCACTTACTTTGCGCTGATGGATGGAAAGGTGTATGACATCACAGAGTGGGC 1888
Qy    601 TGGATGCCAGCCGTGTAGGTATCTCCCCAGATACCCCAGATATCCTCCTATCACATCATT  660
Db   1889 TGGATGCCAGCCGTGTGGGTATCTCCCCAGATACTCCCACAGAGTTCCTTACCACATCATT 1948
Qy    661 TGGTTCTCGGATTCTTCAGGATCGTGAGTCGTGGGTATCTCCCCAGGGGCCGCAGATGCCCCAATGG 720
Db   1949 TGGTTCTCGGGTACCCGGCCAGTGCCGGTACCTCCGGGGGTCCGATGTCCCCGATGTCCCCTGC 2008
Qy    721 TGATCTTCAGGATTTCTTGAGTCGGATCTTTCAAGTCGCAGCCGTGTGCCCCTCAAGTGCCAATGG 780
Db   2009 TGACCTGCAGATTTCTTGAGCCGGATTCTTTCAAGTACCTCCGGGGCCGATGTCCAATGG 2068
Qy    781 GAACTTCTTTGCAGCTCCTCAGCTGCCCTGGAGCCGCGTGCAGCCTCTAAGCCAACAG  840
Db   2069 GAACTTCTTTGCCCACCTCACCCTGGGGACCCACTTCGACCTCTCAGGCCTCTAGGCCCAACAG 2128
Qy    841 CACAGTACCCAAGGGAGAAGCCAAACCTAGCGCGGGAAGAAAGTGAGGAGGCCCTTCCA  900
Db   2129 TTCAGTACCCAAGGGAGAAGCCAAACCTAAACGGCGGGAAGAAAGTGAGGCGGCGCCCTTTCA 2188
Qy    901 ACGTTGATGCCCCTTCTCT- - -TTCCTCAAATCAATGTCAGGGAGTCAAAAGGGCTGT- -  955
Db   2189 ACGTTGACACCCCTTCTTCTTCCTTCCTCAAATCAATGTCAGGGAGTCAAAAGGGCGTGT 2248
Qy    956 - -AGCACAGGATGGAGTTTGATTTAT  979
Db   2249 ACAGCACAGGATGGAATTTGATTTAT 2274
```

RESULT 6
```
BU956189
LOCUS       BU956189         918 bp    mRNA    linear   EST 21-OCT-2002
DEFINITION  AGENCOURT_10613510 NIH_MGC_107 Homo sapiens cDNA clone
            IMAGE:6729715 5', mRNA sequence.
ACCESSION   BU956189
VERSION     BU956189.1  GI:24185761
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 918)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM3053  row: f  column: 18
            High quality sequence stop: 544.
FEATURES             Location/Qualifiers
     source          1..918
                     /organism="Homo sapiens"
```

```
Qy    1014  TTCTTTTTATTTTTGAG-ATGGAGTCTTGCTC-TGTTGCCCAGCTGGAGT 1062
             ||| ||||||||||||| ||  ||||||||||| |||||||| |||||
Db     781  TTCCTTTTATTTTTTGAGAAGGGAGTCTTGCTCTGTTGCCCACCTGGAAT 831
```

RESULT 7
AK045345
LOCUS   AK045345   3022 bp   mRNA   linear   HTC 05-DEC-2002
DEFINITION   Mus musculus adult male corpora quadrigemina cDNA, RIKEN
             full-length enriched library, clone:B230104C23 product:DOPAMINE
             RECEPTOR INTERACTING PROTEIN homolog [Rattus norvegicus], full
             insert sequence.
ACCESSION   AK045345
VERSION   AK045345.1 GI:26337270
KEYWORDS   HTC; CAP trapper.
SOURCE   Mus musculus (house mouse)
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE   High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
             Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE   Normalization and subtraction of cap-trapper-selected cDNAs to
             prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
             Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,T., Itoh,M.,
             Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
             Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
             Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
             Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
             Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE   RIKEN integrated sequence analysis (RISA) system--384-format
             sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
             Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
             Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
             Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
             Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
             Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
             Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
             Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
             Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
             Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
             Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
             Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
             Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
             Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
             Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
             Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
             Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
             Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
             and Hayashizaki,Y.
  TITLE   Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409 (6821), 685-690 (2001)
  MEDLINE   21085660
   PUBMED   11217851
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
             Group Phase I & II Team.
  TITLE   Analysis of the mouse transcriptome based on functional annotation
             of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6 (bases 1 to 3022)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
             Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
             Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
             Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
             Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
             Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
             Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
             Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
             Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
             Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
             Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
             Muramatsu,M. and Hayashizaki,Y.
  TITLE   Direct Submission
  JOURNAL   Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
             Physical and Chemical Research (RIKEN), Laboratory for Genome
             Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
             RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
             Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
             URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
             Fax:81-45-503-9216)
COMMENT   cDNA library was prepared and sequenced in Mouse Genome
             Encyclopedia Project of Genome Exploration Research Group in Riken
             Genomic Sciences Center and Genome Science Laboratory in RIKEN.
             Division of Experimental Animal Research in Riken contributed to
             prepare mouse tissues.
             Please visit our web site for further details.
             URL:http://genome.gsc.riken.go.jp/.
             URL:http://fantom.gsc.riken.go.jp/.
FEATURES   Location/Qualifiers
     source   1..3022
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:B230104C23"
                     /db_xref="taxon:10090"
                     /clone="B230104C23"
                     /sex="male"
                     /tissue_type="corpora quadrigemina"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     CDS   140..2176
                     /note="unnamed protein product; DOPAMINE RECEPTOR
                     INTERACTING PROTEIN homolog [Rattus norvegicus]
                     (SPTR|Q92SG7, evidence: FASTY, 92.7%ID, 100%length.
                     match=2109)"
                     /codon_start=1
                     /putative"
                     /protein_id="BAC32320.1"
                     /db_xref="GI:26337271"
                     /translation="MAQKHPGERRLCGAHRSGGTSLSTSGSSVDPEILSFSGLRDSAE
                     TAPNGTRCLKEHSGPKYTQPPNPAHWSDPSHGPPRGPGPPRGGYPDESETGSEESGV
                     DQELSRENETGYOEDGSPSFLSIPSACNCQGSPGVPEGTYSEEGDGSSSLCHHCTSP
                     ALGEDEELEEYDDEEPLKFPSDFSRVSSGKKPLSRRQKHRFLIKEDVRDSGRREPKA
                     PGRHRLARKRSQTDKRRGLGLWGVEELCQLGAGFWHLIELLVLVGEYVETCGHLIYA
                     CRKLKGSDLDLFPVWVGVWARRLGGWARMMFQFLSQGSHEVCVGLLIRILRVVGAFLLL
                     ALALFLGCLQLGWRFSVGLGNRLGWRDKTAWLFSWLGSPALHHCLTLLKDSRPWQQLV
                     RLIQWGWQELPWVKQRTKKQGNAPVASGRYCQPEEEVTRLLTMAGVPEDELNPFHVLG
                     VEATASDTELKKAYRQLAVMVHPDKNHHPRAEEAFKLRAAWDIVSNPERRKEYEMKR
                     MAENELSRSVNEFLSKLQDDLKEAMNTMMCSRCQGKHRRFEMDREPKSARYCAECNRL
                     HPAEEGDFWAESSMLGLKITYPALMDGKVYDITEWAGCQRVGISPDTHRVPYHISFGS
                     RVPGTSGRQRATPESPPADLQDFLSRIFQVPSGADVQWELLCRTSPWPWDHFDL"
BASE COUNT   753 a   716 c   826 g   727 t
ORIGIN
Query Match   68.6%;   Score 787.4; DB 11; Length 3022;
Best Local Similarity   89.5%;   Pred. No. 1.3e-29;
Matches 883; Conservative   0; Mismatches 96; Indels   8; Gaps   3;

Qy   1 GCCTTGGGTCAAGCAGAATATTAATAGGCAGGGGAATGCACCTGTAGCTAGTGGCGCTA 60
```

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

TITLE

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

TITLE

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

TITLE

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

TITLE

JOURNAL
REFERENCE
AUTHORS

INTERACTING PROTEIN homolog [Rattus norvegicus], full insert
sequence.
AK083878
AK083878.1  GI:26101555
HTC; CAP trapper.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
10349636
2
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
11042159
3
Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861
4
Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,H., Kasukawa,I.,
Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
and Hayashizaki,Y.
Functional annotation of a full-length mouse cDNA collection
Nature 409 (6821), 685-690 (2001)
21085660
11217851
5
The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6  (bases 1 to 4053)
Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,T., Kasukawa,T.,
Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,

[Sequence alignment section — Db/Qy rows]

Db     1345  ... GCCCTGGGTCAAACAGAGAGGACTAAGAAACAGGGCAATGCACCTGTAGCTAGCGGGCGATA  1404
Qy       61  CTGCCAGCCTGAAGAGGAAGTGGCTCGACTCTTGACCATGGCTGGGGTTCCTGAGGATGA    120
Db     1405  CTGCCAGCCTGAAGAGGAAGTGACCCGACTCTTGACCATGGCTGGGGTTCCTGAAGATGA  1464
Qy      121  GCTAAACCCTTCCATGTACTGGGGGTGAGGCCACAGCATCAGATGTTGAACTGAAGAA    180
Db     1465  ACTAAACCCTTTCATGTCTCGGGGGTTGAAGCTACAGCATCCCACACTGAACTAAAGAGA  1524
Qy      181  GGCCTATAGACAGCTGGCAGTGAGTGGTTCATCCTGACAAAATCATCATCCCGGCTGA    240
Db     1525  GGGCCTATAGGCAGCTAGCAGTAATGGTCCATCCTGATAAAAATCACCATCCCGGCTGA  1584
Qy      241  GGAGGCCTTCAAGGTTTGCGAGCAGCTGGGACATTGTCAGCAATGCTGAAAAGCGAAA    300
Db     1585  GGAGGGCCTTCAAAATTTGCGGGCAGCTGGGACATTGTCAGCAACCCAGAGAGGCGGAA  1644
Qy      301  GGAGTAGAGATGAAAACGAATGGCAGAGATGAGCTGAGCCGGTCAGTAAATGAGTTTCT    360
Db     1645  GGAATATGATGAGAAACGATGGCAGAGATGAGCTCAGCCGGTCAGTGAATGAGTTTCT  1704
Qy      361  GTCCAAGCTGCAAGATGACCTCAAGGAGGAGCAATGAATACTATGATGTGTAGCCGATGCCA    420
Db     1705  GTCGAAACTACAAGATGACCTCAAGGAGGAGCAATGAACACGATGATGTGCAGCAGATGCCA  1764
Qgi     421  AGGAAAGCATAGGAGGTTTGAAATGGACCGGGAACCTAAGAGTGCCCAGATACTGTGCTGA    480
Db     1765  AGGAAAGCATAGGAGGTTTGAAATGGACCGGGAACCCCAGAGTGCCCAGATACTGTGCTGA  1824
Qy      481  GTGTAATAGGCTGCATCCTGCTGAGGAAGGAGACTTTTGGGCAGAGTCAAGCATGTTGGG    540
Db     1825  GTGTAACAGGCTGCATCCTGCTGAGGAAGGAGACTTTTGGGCAGAGTCGAGCATGCTGGG  1884
Qy      541  CCTCAAGATCACCTACTTTGCACTGATGATGATGGAAAGGTGTATGACATCACAGAGTGGGC    600
Db     1885  CCTCAAGATCACTTACTTTGCGCTGATGATGGAAAGGTGTATGACATCACAGAGTGGGC  1944
Qy      601  TGGATGCCAGCGTGTAGGTATCTCCCCAGATATCCCCTATCACATCTCATT    660
Db     1945  TGGATGCCAGCGTGTGGGTATCTCCCCAGATACTCACAGAGTTCCTTACCACATCTCATT  2004
Qy      661  TGGTTCTCGGGATTCCAGGCACCAGAGGGCGCAGAGAGCCACCCCAGATGCCCCTCCTGC    720
Db     2005  TGGTTCTCGGGTACCGGCACCAGTGCCGCGGCCAGTGGGCCACTCCAGAGTCCCCTCCTGC  2064
Qy      721  TGATCTTCAGGATTCTTGAGTCGGATCTTTCAAGTACCC--CCAGGGCGACGATGCCCAATG    779
Db     2065  TGACCTGCAGGATTTCTTGAGCCGGATCTTTCAAGTACCCTCCCGGGGCCCGATGTCCAATG  2124
Qy      780  GGAACTTCTTGCAGCTCCTCAGCGTGCCCCTGCCCCTGAGCCGGCTCTCAAGCCCCACA    839
Db     2125  GGAACTTCTTTGCCGCACCTCACCCTGGCACCTCGACCTCTCAGGCCCCCACA  2184
Qy      840  GCACAGTACCCAGGGAGAGCCAAACCTCAAGCCGGGAGAAGAAAGTGAGGAGGCCCTCC    899
Db     2185  GTTCAGTACCCAAGGGAGAGGAGCCAAACCTAAACGCGGAGAAGAAAGTGAGGCGGCCCTTTC  2244
Qy      900  AAGTTGATGCCCCTTCT---TTCCTCAAATCAAATCAATGTCAGGGAGTCAAAAGGGCTGT-    955
Db     2245  AACGACAGACACCCCCTTCTCTTCCTTCCTCAAATCAATGTCAGGGAGTCAAAAGGGCTGTG  2304
Qy      956  ---AGCCAGGGATGGAGTTGATTTAT    979
Db     2305  TACAGCACAGGGATGGAATTTGATTTAT  2331

RESULT 8
AK083878
LOCUS       AK083878                4053 bp    mRNA    linear    HTC 05-DEC-2002
DEFINITION  Mus musculus 12 days embryo spinal ganglion cDNA, RIKEN full-length
            enriched library, clone:D130038P18 product:DOPAMINE RECEPTOR

```
Qy   421  AGGAAAGCATAGGAGAGGTTTGAAATGGACCGGGAACCTAAGAGTGCCAGATACTGTCTGA  480
          |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db  1720  AGGAAAGCATAGGAGAGGTTTGAAATGGACCGGGAACCCAAGAGTGCCAGATACTGTCTGA  1779

Qy   481  GTGTAATAGGTGCATCCTGCTGAGGAAGGAGACTTTTGGGCAGAGTCAAGACATGTGGG    540
          ||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db  1780  GTGTAACAGGCTGCATCCTGCTGAGGAAGGAGAGAGACTTTTGGGCAGAGTCGAGCATGCTGGG  1839

Qy   541  CCTCAAGATCACCTACTTGCACTGATGGATGGATGGATGAAAGGTGTATGACATCACAGAGTGGGC  600
          |||||||||||||||||||||||||||| ||||| |||||||||||||||||||||||||
Db  1840  CCTCAAGATCACTTACTTTGGGCTGATGGATGGAAAGGTGTATGACATCACAGAGTGGGC  1899

Qy   601  TGGATGCCAGCGTGTAGGTATCTCCCAGATACCCACGAGTCCCTATCACATCTCATT  660
          |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db  1900  TGGATGCCAGCGTGTGGGTATCTCCCAGATACTCACAGAGTTCCTTACCACATCTCATT  1959

Qy   661  TGGTTCTCGGATTCCAGCACCAGAGGGCGGCAGAGAGCCACCCCAGATGCCCCTCCTGC  720
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db  1960  TGGTTCTCGGGATACCCGGCACCAGTGCCGCGGCAGGGGCCACTCCAGA-GTCCCTCCTGC  2018

Qy   721  TGATCTTCAGGATTCTTGAGTCGGATTCTTGAGCCGGATCTTTCAAGTACCCCAGGGGCAGATGCCCAATGG  780
          ||| ||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db  2019  TGACCTGCAGGATTCTTGAGCCGGATCTTTCAAGTACCTCCGGGGCCGATGTCCAATGG  2078

Qy   781  GAACTTCTTTGCAGCTCCTCAGCTGCCTGCCTGCGTGAGCCGCTCTAAGCCACACAG  840
          ||||||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db  2079  GAACTTCTTTGCCGCACCCTGCCCTGGACCCTGGACCTTCGACCTTCTAGGCCCAACAG  2138

Qy   841  CACAGTACCCAAGGAGAAGCCAAACCTAAGCGGCGGGAAGAAAGTGAGGAGGCCCTTCCA  900
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db  2139  TTCAGTACCCAAGGAGAAGCCAAACCTAAACGGCGGGAAGAAAGTGAGGGGCGCCCTTTCA  2198

Qy   901  ACGTTGATGCCCCTTCT---TTCTCAAATCAATGTCAGGGAGTCAAAAGGGCTGT--  955
          ||||||||||||| |||   |||| |||||||||||||||||||||||||||||||
Db  2199  ACGATGACACCCCTTCTTCTTCCTTCCAAATCAATGTCAGGGAGTCAAAAGGGCTGTGT  2258

Qy   956  --AGCACAGGATGGAGTTTGATTAT  979
            ||||||||||||||||||||||||
Db  2259  ACAGCACAGGATGGGAGATTTGATTTAT  2284
```

RESULT 9
BM802135
LOCUS   BM802135 AGENCOURT_6461654 NIH_MGC_88 Homo sapiens   957 bp   mRNA   linear   EST 05-MAR-2002
DEFINITION AGENCOURT_6461654 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:5559503
            5', mRNA sequence.
ACCESSION   BM802135
VERSION   BM802135.1 GI:19118958
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 957)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12284 row: c column: 24
            High quality sequence stop: 621.
FEATURES   Location/Qualifiers
     source   1..957
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"

Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Tagami,M., Tagawa,K., Takahashi,F., Takaku-Akahira,S.,
Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
Muramatsu,M. and Hayashizaki,Y.
TITLE   Direct Submission
JOURNAL   Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT   cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.go.jp/
            URL:http://fantom.gsc.riken.go.jp/.
FEATURES   Location/Qualifiers
     source   1..4053
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:D130038P18"
                     /db_xref="taxon:10090"
                     /clone="D130038P18"
                     /tissue_type="spinal ganglion"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="12 days embryo"
     misc_feature   94..2205
                     /note="DOPAMINE RECEPTOR INTERACTING PROTEIN homolog
                     [Rattus norvegicus] (SPTR|Q925G7, evidence: FASTY,
                     92.7%ID, 100%length, match=2109)
                     putative"
BASE COUNT   1060 a   903 c   1039 g   1051 t
ORIGIN

Query Match   68.6%; Score 786.4; DB 11; Length 4053;
Best Local Similarity   89.5%; Pred. No. 1.2e-29;
Matches 882; Conservative 0; Mismatches 96; Indels 8; Gaps 3;

```
Qy     1  GCCTTGGGTCAAGCAGAGAATATTAATAGGCAGGGGAATGCACCTGAGCTAGTGGGCGCTA    60
          ||| ||||||| |||| |||| |||| ||||||||||||||||||||||||||||||||||
Db  1300  GCCCTGGGTCAAACAGAGAGACTAAGAAACAGGGCAATCAGGGCAATGCACCTGAGCTAGTCGGCGGCGATA  1359

Qy    61  CTGCCAGCCTGAAGAGAGGAAGTGGCTCGACTCTTGACCGACTCTTGACCGGTTCCTGAGGATGA   120
          ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db  1360  CTGCCAGCCTGAAGAGAGGAAGTGGCGAAGTGACCGACTCTTGACCGGTTCCTGAAGATGA  1419

Qy   121  GCTAAACCCTTCCATGTACTGGGGGTTGAGCGCCACAGATCAGATGTTGAACTGAAGAA   180
          |||| ||||||||||||| |||||||||||||||||||||||| |||||||| ||||||
Db  1420  ACTAAACCCTTTCATGTGCTGGGGGTTGAAGCTACAGCATCCGACACTGAACTAAAGAA  1479

Qy   181  GGCCTATAGACAGCTGCCAGTGATGGTTCATCCTGACAAAAATCATCATCCCCGGGCTGA   240
          |||||||||| |||||||||||| ||||||||||| ||||||||||||||||||||||||
Db  1480  GGCCTATAGGCAGCTAGCAGTAATGGTCCATCCTGATAAAAATCACCATCCCCGGGCTGA  1539

Qy   241  GGAGCGCCTTCAAGGTTTTGCGAGCAGCTGGGACATTGTCGCAATGCTGAAAAGCGAAA   300
          ||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db  1540  GGAGGCCCTTCAAAATTTTGCGGGGCAGCTTGGGACATTGTCAGCAACCCCAGAGAGCGGGAA  1599

Qy   301  GGAGTATAGAGATGAAACGAATGGCAGAGAAATGGAGCTGAGCCGGTCAGTAAATGAGTTTCT   360
          ||| |||||| |||||||||||||||||||||||||||||||| ||||| |||||||||||
Db  1600  GGAATATGAGATGAAACGGATGGGCAGAGAAATGAGCTCAGCCGGTCAGTGAATGAGTTTCT  1659

Qy   361  GTCCAAGTGCAAGATGACCTCAAGGAGGCAATGAATACTATGAATGTGTAGCCCGATGCCA   420
          |||| |||||| ||||||||||||||||||||||||| |||| |||||||||||||||||||
Db  1660  GTCGAAACTACAAGATGACCTCAAGGAGGCAATGAACACCGATGAGTGTGCAGCAGATGCCA  1719
```

RESULT 10

BG763500

LOCUS   BG763500   6027355591F1 NIH_MGC_49 Homo sapiens cDNA clone IMAGE:4860694 5',
             878 bp   mRNA   linear   EST 15-MAY-2001
             mRNA sequence.
DEFINITION   BG763500 6027355591F1 NIH_MGC_49 Homo sapiens cDNA clone IMAGE:4860694 5',
ACCESSION   BG763500
VERSION   BG763500.1 GI:14074153
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 878)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished
COMMENT   Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: ATCC/DCTD/DTP
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLCM1718 row: f column: 23
             High quality sequence stop: 811.

FEATURES   Location/Qualifiers
     source   1..878
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4860694"
                     /tissue_type="melanotic melanoma, high MDR (cell line)"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_49"
                     /note="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Size-selected >500bp for average insert size
                     1.8kb. Library constructed by Ling Hong in the laboratory
                     of Gerald M. Rubin (University of California, Berkeley)
                     using ZAP-cDNA synthesis kit (Stratagene) and Superscript
                     II RT (Life Technologies). Note: this is a NIH_MGC
                     Library."

BASE COUNT   227 a   211 c   238 g   201 t   1 others
ORIGIN

Query Match   66.2%;   Score 759.4; DB 12; Length 878;
Best Local Similarity 95.1%;   Pred. No. 6.1e-28;
Matches 837; Conservative 0; Mismatches 37; Indels 6; Gaps 5;

Qy   263 GCAGCTTGGGACATTGTCAGCAATGCTGAAAAGCGAAAAGGAGTATGAGATGAAAACGAATG 322
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2 GCAGCTTGGGACATTGTCAGCAATGCTGAAAAGCGAAAAGGAGTATGAGATGAAAACGAATG 61

Qy   323 GCAGAGAATGAGCTGAGCCGGTCAGTAAATGAGTTTCGTCGTGAAGGAGATGACCTC 382
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   62 GCAGAGAATGAGCTGAGCCGGTCAGTAAATGAGTTTCGTCGTGAAGGAGATGACCTC 121

Qy   383 AAGGAGGCAATGAATACTATGATGTGTAGCCGATGCCAAGGAAAAGCATAGGAGGTTTGAA 442
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   122 AAGGAGGCAATGAATACTATGATGTGTAGCCGATGCCAAGGAAAAGCATAGGAGGTTTGAA 181

Qy   443 ATGGACCGGGAACCTAAGAGTGCCAGAGTACTGTGCGAGTGTAATAGGCTGCATCCTGCT 502
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   182 ATGGACCGGGAACCTAAGAGTGCCAGAGTACTGTGCGAGTGTAATAGGCTGCATCCTGCT 241

Qy   503 GAGGAAGGAGACTTTGGGCAGAGTCAAGCATGTTGGGCCTCAAGATCACCTACTTTGCA 562
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   242 GAGGAAGGAGACTTTGGGCAGAGTCAAGCATGTTGGGCCTCAAGATCACCTACTTTGCA 301

Qy   563 CTGATGGATGGAAAGTGTATGACATCACAGAGTGGCTGGATGCCGCAGCGTGTAGGTATC 622

             /clone="IMAGE:5559503"
             /tissue_type="duodenal adenocarcinoma, cell line"
             /lab_host="DH10B (phage-resistant)"
             /clone_lib="NIH_MGC_88"
             /note="Organ: small intestine; Vector: pCMV-SPORT6;
             Site 1: NotI; Site 2: SalI; Cloned unidirectionally.
             oligo-dT primed. Average insert size 1.767 kb. Library
             enriched for full-length clones and constructed by Life
             Technologies. Note: this is a NIH_MGC Library."

BASE COUNT   243 a   240 c   246 g   224 t   4 others
ORIGIN

Query Match   66.7%;   Score 764.6; DB 12; Length 957;
Best Local Similarity 97.9%;   Pred. No. 3.3e-28;
Matches 796; Conservative 0; Mismatches 14; Indels 3; Gaps 2;

Qy   219 AAAATCATCATCCCCGGGCTGAGGAGGCCTTCAAGGTTTGCGAGCAGCTTGGGACATTG 278
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10 AAAATCATCATCCCCGGGCTGAGGAGGCCTTCAAGGTTTGCGAGCAGCTTGGGACATTG 69

Qy   279 TCAGCAATGCTGAAAAGCGAAAAGGAGTATGAGATGAAAACGAATGCCAGAGAATGAGCTGA 338
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   70 TCAGCAATGCTGAAAAGCGAAAAGGAGTAGTAGATGAAAACGAATGCCAGAGAATGAGCTGA 129

Qy   339 GCCGGTCAGTAAATGAGTTTCGTCCAAGCTGCAAGATGACCTCAAGGAGGCAATGAATA 398
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   130 GCCGGTCAGTAAATGAGTTTCGTCCAAGCTGCAAGATGACCTCAAGGAGGCAATGAATA 189

Qy   399 CTATGATGTGTAGCCGATGCCAAGGAAAAGCATAGGAGGTTTGAAATGGACCGGGAACCTA 458
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   190 CTATGATGTGTAGCCGATGCCAAGGAAAAGCATAGGAGGTTTGAAATGGACCGGGAACCTA 249

Qy   459 AGAGTGCCAGATACTGTGCGTGAGTGTAATAGGCTGCTGCATCCTGCTGAGGAAGGAGACTTTT 518
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   250 AGAGTGCCAGATACTGTGCGTGAGTGTAATAGGCTGCTGCATCCTGCTGAGGAAGGAGACTTTT 309

Qy   519 GGGCAGAGTCAAGCATGTTGGGCCTCAAGATCACCTACTTTGCACTGATGATGGAAAGG 578
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   310 GGGCAGAGTCAAGCATGTTGGGCCTCAAGATCACCTACTTTGCACTGATGATGGAAAGG 369

Qy   579 TGTATGACATCACAGAGTGGCTGGATGCCAGCGTGTAGTAGTATCCCCAGATACCCACA 638
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   370 TGTATGACATCACAGAGTGGCTGGATGCCAGCGTGTAGTAGTATCCCCAGATACCCACA 429

Qy   639 GAGTCCCCATCACATCCATTGGTTCTCGGATTCCAGGACCAGAGGGCGGCAGAGAG 698
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   430 GAGTCCCCATCACATCCATTGGTTCTCGGATTCCAGGACCAGAGGGCGGCAGAGAG 489

Qy   699 CCACCCCAGATGCCCCCTCCTGCTGATCTTCAGGATTCTTCAGGTCGGATCTTCAAGTAC 758
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   490 CCACCCCAGATGCCCCCTCCTGCTGATCTTCAGGATTCTTCAGGTCGGATCTTCAAGTAC 549

Qy   759 CCCCAGGGGCAGATGCCCAATGGGAACTTCTTTGCAGCTCCCAGCCTCCTGCCCTGGAGCCG 818
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   550 CCCCAGGGGCAGATGCCCAATGGGAACTTCTTTGCAGCTCCCAGCCTCCTGCCCTGGAGCCG 609

Qy   819 CTGCAGCCTCTAAGCCAACAGCACAGTACCCAAGGAGAAGCCAAACCTAAGCGGCGGGA 878
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   610 CTGCAGCCTCTAAGCCAACAGCACAGTACCCAAGGAGAAGCCAAACCTAAGCGGCGGGA 669

Qy   879 AGAAAGTGAGGAGGCCCTTCCAACGTGATGCCCCCTTCTCTTCCTCAAATCAATGTCAG 938
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   670 AGAAAGTGAGGAGGCCCTTCCAACGTGATGCCCCCTTCTCTTCCTCAAATCAATGTCAG 729

Qy   939 GGAGTCAAAAGGGCTGTAGCACA--GGATCGGAGTTTGATTTAT-CCTCTCCCCCCACA 995
           |||||||||||||||||| ||| |||||||||||||||||| ||||||||||||||||
Db   730 GGAGTCAAAAGGGCTGTAGCACAAGGATGGCACAAGGATGGGAGTTTGATTTATCCCCTCCCTCCCCCCAAAA 789

Qy   996 CCTAGGAACTGAATCTTTTCTTTTATTTTTT 1028
           |||||||||||||||| ||||||||| |||
Db   790 CCTAGGAACTGGAATCTTTTTCCTTTATTTT 822

```
                /tissue_type="adenocarcinoma, cell line"
                /lab_host="DH10B (phage-resistant)"
                /clone_lib="NIH_MGC 107"
                /note="Organ: breast; Vector: pOTB7; Site_1: EcoRI;
                Site_2: XhoI; cDNA made by oligo-dT priming.
                Directionally cloned into EcoRI/XhoI sites using the
                following 5' adaptor: GGCACGAG(G).   Library constructed by
                Ling Hong in the laboratory of Gerald M. Rubin (University
                of California, Berkeley) using ZAP-cDNA synthesis kit
                (Stratagene) and Superscript II RT (Life Technologies).
                Note: this is a NIH_MGC Library."

BASE COUNT      239 a    241 c    261 g    213 t    1 others
ORIGIN

Query Match      64.9%;   Score 744.2;  DB 13;  Length 955;
Best Local Similarity 97.8%;   Pred. No. 2.9e-27;
Matches 797;   Conservative   0;   Mismatches 13;   Indels   5;   Gaps   4;

Qy    182 GCCTATAGACAGCTGCAGTGATGGTTCATCCTGACAAAATCATCATCCCGGGCTGAG 241
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GCCTATAGACAGCTGCAGTGATGGTTCATCCTGACAAAATCATCATCCCGGGCTGAG  60

Qy    242 GAGGCCTTCAAGGTTTTGGGAGCAGCTTGGGACATTGTCAGCAATGCTGAAAAGGCGAAAG 301
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GAGGCCTTCAAGGTTTTGGGAGCAGCTTGGGACATTGTCAGCAATGCTGAAAAGGCGAAAG 120

Qy    302 GAGTATGAGATGAGAAACGAATGGCACAGAGAATGAGCTGAGCCGGTCAGTAAATGAGTTTCTG 361
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GAGTATGAGATGAGAAACGAATGGCACAGAGAATGAGCTGAGCCGGTCAGTAAATGAGTTTCTG 180

Qy    362 TCCAAGCTGCAAGATGACCTCAAGGAGCAATGATCATGATGTGTAGGCGATCCAA 421
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TCCAAGCTGCAAGATGACCTCAAGGAGCAATGATCATGATGTGTAGGCGATCCAA 240

Qy    422 GGAAAGCATAGGAGGTTGAAATGACCGGGAACCTAAGAGTGACCAAGAGTCTTTTGGGCAGACTTTTGGGCAGCATGTGGGC 481
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GGAAAGCATAGGAGGTTGAAATGACCGGGAACCTAAGAGTGACCAAGAGTCTTTTGGGCAGACTTTTGGGCAGCATGTGGGC 300

Qy    482 TGTAATAGGCTGCATCCTGCTGAGGAAGGAGACTTTTGGGCAGAGACTTTTGGGC 541
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TGTAATAGGCTGCATCCTGCTGAGGAAGGAGACTTTTGGGCAGAGACTTTTGGGC 360

Qy    542 CTCAAGATCACCTACTTTGCACTGATGATGGAAAGGTGTATGACATCACAGAGTGGCT 601
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CTCAAGATCACCTACTTTGCACTGATGATGGAAAGGTGTATGACATCACAGAGTGGCT 420

Qy    602 GGATGCCAGCGTGTAGGTATCTCCCAGATACCCAGAGTCCCCTATCACATCTCATT 661
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GGATGCCAGCGTGTAGGTATCTCCCAGATACCCAGAGTCCCCTATCACATCTCATT 480

Qy    662 GGTTCTCGGATTCCAGGCACCAGAGGGCCGCAGAGAGCCACCCCGAGATGCCCCTCCTGCT 721
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GGTTCTCGGATTCCAGGCACCAGAGGGCCGCAGAGAGCCACCCCGAGATGCCCCTCCTGCT 540

Qy    722 GATCTTCAGGATTCTTGAGTCGATCTTTCAAGTACCCCA-GGGCAGATGCCCAATGG 780
          |||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db    541 GATCTTCAGGATTCTTGAGTCGATCTTTCAAGTACCCCTAAGTGGCAGATGCCCAATGG 600

Qy    781 GAACTTCTTTGCAGTCCTCAGCCTCGCCCCTGGAGCCGCTGCAGCCTCTAAGCCCAACAG 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 GAACTTCTTTGCAGTCCTCAGCCTCGCCCCTGCCCCTGGAGCCGCTGCAGCCTCTAAG-CCAACAG 659

Qy    841 CACAGTACCCAAGGGAGAAGCCAAACCTAAGCGGCGGAAGAAGTGAGAGGCCCTTCCA 900
          ||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db    660 CACAGTACCCAAGGGAGAAGCCAAACCTAAGCGGCGGAAGAAGTGAGAGGCCCTTCCA 719

Qy    901 ACGTTGATGCCCCTTCTCTTTCCTCAAATCAATGTCAGGGAGTCAAAAGGGCTGTAGCAC 960
          |||||||||||||||||||||||||||| |||||||||||||||||||||||
Db    720 CCGTTGATGCCCCTTCTCTTTCCTC-AATCAATGCAGGGAGTCAAAAGGGCTCAAAAGGCTGTAACAC 778

Qy    961 AGG--ATGGAGTTTGATTTATCCTCCTCCCCCAA 993
          |||   |||||||||||||||||||||||||||||
Db    779 AGGGATGGAGATTTGATTTATCCCCCTCCCCCCA 813
```

```
Db    302 CTGATGGATGGAAAGGTGTATGACATCACAGAGTGGGCTGGATGGTAGGTAGGTATC 361
          |||||||||||||||||||||||||||||||||||||||||||||||
Qy    623 TCCCCAGATACCCACCAGAGTCCCTATCACATCTCACATTGGTTCTCGGATTCTTCCAGGCACC 682
          |||||||||||||||||||||||||||||||||||||||||||||||
Db    362 TCCCCAGATACCCACCAGAGTCCCTATCACATCTCACATTGGTTCTCGGATTCTTCCAGGCACC 421
          |||||||||||||||||||||||||||||||||||||||||||||||
Qy    683 AGAGGGGCGCAGAGAGCCACCCCGAGATGCCCCTCCTGCTGATCTTCAGGATTCTTCGAGT 742
          |||||||||||||||||||||||||||||||||||||||||||||||
Db    422 AGAGGGGCGCAGAGAGCCACCCCGAGATGCCCCTCCTGCCCATGGCCCGCTGATCTTCAGGATTCTTCGAGT 481
          |||||||||||||||||||||||||||||||||||||||||||||||
Qy    743 CGATCTTCAAGTACCCCCAGGGCAGATGCCCAATGGGAACTTCTTTGCAGTCCTCAG 802
          |||||||||||||||||||||||||||||||||||||||||||||||
Db    482 CGATCTTCAAGTACCCCCAGGGCAGATGCCCAATGGGAACTTCTTTGCAGTCCTCAG 541
          |||||||||||||||||||||||||||||||||||||||||||||||
Qy    803 CCTGCCCCTGGAGCCGCTGCAGCCTCTAAGCCCAACAGCACAGTACCCAAGGGAGAAGCC 862
          |||||||||||||||||||||||||||||||||||||||||||||||
Db    542 CCTGCCCCTGGAGCCGCTGCAGCCTCTAAGCCCAACAGCACAGTACCCAAGGGAGAAGCC 601
          |||||||||||||||||||||||||||||||||||||||||||||||
Qy    863 AAACCTAAGCGGCGGCGGGAGAAGTGAGAGGCCCTTCCAACGTTGATGCCCCTTCTTTC 922
          |||||||||||||||||||||||||||||||||||||||||||||||
Db    602 AAANCTAAGCGGCGGCGGGAGAAGTGAGAGGCCCTTCCAACGTTGATGCCCCTTCTTTC 661
          |||||||||||||||||||||||||||||||||||||||||||||||
Qy    923 CTCAAATCAATGTCAGGGAGTC-AAAAGGGCTGTAGCACAGGATGGAGTTTGATTTATCC 981
          |||||||||||||||||||||||||||||||||||||||||||||||
Db    662 CTCAAATCAATGTCAGGGAGTCAAAAGGGCTGTAGCACAGGATGGAG-TTGATTTATCC 720
          |||||||||||||||||||||||||||||||||||||||||||||||
Qy    982 CTCTCCCCAACACCTAGGAACTGAATCTTTTTTATTTTTTGAGATGGAGTCTT 1041
          |||||||||||||||||||||||||||||||||||||||||||||||
Db    721 CTCCTCCCCAA-AACTAGGAACTGAATCTTTGCTTGTATTTTTGAGATGGAGTCTTGC 779
          |||||||||||||||||||||||||||||||||||||||||||||||
Qy   1042 GCTCTGTTGCCCAGCTGAGTGC-AGTGGTGTGATCTCAGCTTACTGCAACCTCTGTCTC 1100
          |||||||||||||||||||||||||||||||||||||||||||||||
Db    780 TCTTGTTGTTCCCAGGCTGGAGTGCAAGTGGTGTGATCTCAGCTTACTGCAACCTCTGTCTC 839
          |||||||||||||||||||||||||||||||||||||||||||||||
Qy   1101 CCGGGTTCAAGCAATTCTCCCATCTCAGCCTCTCGAGTAG 1140
          |||||||||||||||||||||||||||||||||||||||||||||||
Db    840 CCGGGTTCAGGCAAT--CTCCATCTCAAGCTCCTGAGTAG 877
```

```
RESULT 11
BU538351
LOCUS       BU538351                 955 bp    mRNA    linear    EST 13-SEP-2002
DEFINITION  AGENCOURT 10181238 NIH_MGC_107 Homo sapiens cDNA clone
            IMAGE:6568891 5', mRNA sequence.
ACCESSION   BU538351
VERSION     BU538351.1  GI:22848792
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 955)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2755 row: e  column: 19
            High quality sequence stop: 583.
FEATURES             Location/Qualifiers
     source          1..955
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6568891"
```

RESULT 12
BI333110
LOCUS   BI333110   785 bp   mRNA   linear   EST 30-JUL-2001
DEFINITION 602996387F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:5138876 5',
            mRNA sequence.
ACCESSION   BI333110
VERSION   BI333110.1 GI:15017767
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 785)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: Incyte Genomics, Inc.
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM1341 row: m column: 21
            High quality sequence stop: 775.
FEATURES   Location/Qualifiers
     source   1..785
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5138876"
                     /tissue_type="cervical carcinoma cell line"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_12"
                     /note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.4 kb. Library prepared by Life
                     Technologies."
BASE COUNT   192 a   202 c   202 g   189 t
ORIGIN

Query Match   61.6%; Score 706.4; DB 12; Length 785;
Best Local Similarity   98.4%; Pred. No. 1.9e-25;
Matches 777; Conservative   0; Mismatches   6; Indels   7; Gaps   6;

Qy   337 GAGCCGGTCAGTAAATGAGTTTCTGTCCAAGCTGCAAGATGACCTCAAGGAGGCAATGAA 396
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GAGCCGGTCAGTAAATGAGTTTCTGTCCAAGCTGCAAGATGACCTCAAGGAGGCAATGAA 60

Qy   397 TACTATGATGTGTAGCCGATGCCAAGGAAAGCATAGGAGTTGAAATGACCGGGAACC 456
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 TACTATGATGTGTAGCCGATGCCAAGGAAAGCATAGGAGTTGAAATGACCGGGAACC 120

Qy   457 TAAGAGTGCCAGATACTGTGCTGCGAGTGTAATAGGCTGTCATCCTGCTGAGGAAGGAGACTT 516
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 TAAGAGTGCCAGATACTGTGCTGCGAGTGTAATAGGCTGTCATCCTGCTGAGGAAGGAGACTT 180

Qy   517 TTGGGCAGAGTCAAGCATGTGGGCCTCAAGATCACCTACTTTGCACTGATGGATGGAAA 576
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 TTGGGCAGAGTCAAGCATGTGGGCCTCAAGATCACCTACTTTGCACTGATGGATGGAAA 240

Qy   577 GGTGTATGACATCACAGAGTGGGCTGGATGCCCAGGCGTGTAGGTATCTCCCAGATACCCA 636
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GGTGTATGACATCACAGAGTGGGCTGGATGCCCAGGCGTGTAGGTATCTCCCAGATACCCA 300

Qy   637 CAGAGTCCCCTATCACATCATTGGTTCTCGGATTCCAGGCACCAGGGGCGCAGAG 696
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 CAGAGTGCCCTATCACATCATTGGTTCTCGGATTCCAGGCACCAGAGGGCGCAGAG 360

Qy   697 AGCCACCCCAGATGCCCTCCTGCTGATCTTCAGGATTTCTTGAGTCGGATCTTTCAAGT 756
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 AGCCACCCCAGATGCCCTCCTGCTGATCTTCAGGATTTCTTGAGTCGGATCTTTCAAGT 420

Qy   757 ACCCCAGGGCAGATGCCCAATGGGAACTTCTTTGCAGCTCCTCAGCCTGCCCCTGGAGC 816
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 ACCCCCA-GGCAGATGCCCAATGGGAACTTCTTTGCAGCTCCTCAGCCTGCCCCTGGAGC 479

Qy   817 CGCTGAGCCTCTAAGCCCAACAGCACAGTACCCAAGGGAGAAGCCAAACCTAAGCGGCG 876
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   480 CGCTGAGCCTCTAAGCCCAACAGCACAGTACCCAAGGGAGAAGCCAAACCTAAGCGGCG 539

Qy   877 GAAGAA-AGTGAGGAGGAGCCCTTCCAACGTTGATGCCCTTCTTCCTCAAATCAATGT 935
          |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   540 GAAGAACAGTGAGGAGGAGCCCTTCCAACGTTGATGCCCTTCTTCCTCAAATCAATGT 599

Qy   936 CAGGGAGTCAAAAGGCTGTAGCACAGGAGTGGAGTTTGATTTATTATCCTCCCCCAACA 995
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   600 CAGGGAGTCAAAAGGCTGTAGCACAGGAGTGGAGTTTGATTTATTATCCTCCCCCAACA 659

Qy   996 CCTAGGAACTGAATCTTTTCTTTTATTTTTTGAGATGGAGTCTTGCTCTG-TTGCCCA 1054
          |||||||||||||||||||| |||| ||||||||||||||||||||||||
Db   660 CCTAGGAACTGAATCTTT--CTTTTTATATTTGAGATGGAGTCTTGCTCTGTTTGCCCA 717

Qy   1055 GCTGGGAGTGCAGTGGTGTGATCCTCAGCTTACTGCAACCTCTGTTCCCGGGTTCAAGCAA 1114
          |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db   718 GCTGGGAGTGCAGTGGTGTGATCCTCAG-TTACTGCAA-CTC-TGTCTCCCGGGTTCAAGCAA 775

Qy   1115 TTCTCCCATC 1124
          |||||| ||
Db   776 TTCTCCCATTC 785

RESULT 13
BQ691177
LOCUS   BQ691177   904 bp   mRNA   linear   EST 15-JUL-2002
DEFINITION AGENCOURT_8340957 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:6249114
            5', mRNA sequence.
ACCESSION   BQ691177
VERSION   BQ691177.1 GI:21816493
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 904)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2390 row: a column: 19
            High quality sequence stop: 609.
FEATURES   Location/Qualifiers
     source   1..904
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6249114"
                     /tissue_type="ductal carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_110"
                     /note="Organ: pancreas; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Library constructed by

ORGANISM   Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 995)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/,
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished
COMMENT   Contact: Robert Strausberg, Ph.D.
           Email: cgapbs-r@mail.nih.gov
           Tissue Procurement: ATCC/DCTD/DTP
           cDNA Library Preparation: Life Technologies, Inc.
           cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
           DNA Sequencing by: Agencourt Bioscience Corporation
           Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LLAM12200 row: f column: 03
           High quality sequence stop: 638.
FEATURES   Location/Qualifiers
     source   1..995
                    /organism="Homo sapiens"
                    /mol_type="mRNA"
                    /db_xref="taxon:9606"
                    /clone="IMAGE:5526530"
                    /tissue_type="melanotic melanoma"
                    /lab_host="DH10B (phage-resistant)"
                    /clone_lib="NIH_MGC_72"
                    /note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
                    Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                    Average insert size 2 kb. Library constructed by Life
                    Technologies."
BASE COUNT   224 a   273 c   249 g   248 t
ORIGIN

Query Match   58.6%; Score 672.2; DB 12; Length 995;
Best Local Similarity   99.6%; Pred. No. 6.5e-24;
Matches   674; Conservative   0; Mismatches   3; Indels   0; Gaps   0;

Qy   471 ACTGTGCTGAGTGTAATAGGCTGCATCCTGCTGAGGAAGGAGACTTTTGGGCAGAGTCAA 530
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ACTGTGCTGAGTGTAATAGGCTGCATCCTGCTGAGGAAGGAGACTTTTGGGCAGAGTCAA 60

Qy   531 GCATGTGGGCCTCAAGATCACCTACTTTGACTGATGGAAAGGTGTATGACATCA 590
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GCATGTGGGCCTCAAGATCACCTACTTTGACTGATGGAAAGGTGTATGACATCA 120

Qy   591 CAGAGTGGGCTGGATGCGCAGCGTGTAGGTATCTCCCCAGATACCCCAGAGTCCCCTATC 650
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 CAGAGTGGGCTGGATGCGCAGCGTGTAGGTATCTCCCCAGATACCCCAGAGTCCCCTATC 180

Qy   651 ACATCTCATTTGGTTCTCGGATTCCAGGCACCCAGAGGCGCGCAGAGAGCCACCCCAGATG 710
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 ACATCTCATTTGGTTCTCGGATTCCAGGCACCCAGAGGCGCGCAGAGAGCCACCCCAGATG 240

Qy   711 CCCCTCCTGCTGATCTTCAGGATTTCTTGAGTCGGATCTTTCAGGTCGGATCTTTCAGCCTCTA 770
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 CCCCTCCTGCTGATCTTCAGGATTTCTTGAGTCGGATCTTTCAGTCGGATCCTGAGCCTCTA 300

Qy   771 TGCCCAATGGAACTTCTTTGCAGCTCTTGCAGCGTCCTCAGCCTCCTGCGAGCCTGAGCCTCTA 830
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 TGCCCAATGGAACTTCTTTGCAGCTCTTGCAGCGTCCTCAGCCTCCTGCGAGCCTGAGCCTCTA 360

Qy   831 AGCCCAACAGCACAGTACCCAAGGAGGAAGCCAAACCTAAGCGCGGAAGAAAGTGAGGA 890
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 AGCCCAACAGCACAGTACCCAAGGAGGAAGCCAAACCTAAGCGCGGAAGAAAGTGAGGA 420

Qy   891 GGCCCTTCCAACGTTGATGCCCTTCCTTCTTTCTCTTTCCTCAAATCAATTCAGGGAGTCAAAAGG 950
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 GGCCCTTCCAACGTTGATGCCCTTCCTTCTTTCTCTTTCCTCAAATCAATTCAGGGAGTCAAAAGG 480

Qy   951 GCTGTAGCACAGGATGGAGTTTGATTTATCCTCCCCCAACACTTAGGAACTGAATC 1010
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 GCTGTAGCACAGGATGGAGTTTGATTTATCCCTCCCCCAACACCTTAGGAACTGAATC 540

BASE COUNT   219 a   244 c   239 g   202 t
ORIGIN

Query Match   60.4%; Score 692.8; DB 13; Length 904;
Best Local Similarity   95.5%; Pred. No. 7.6e-25;
Matches   745; Conservative   0; Mismatches   32; Indels   3; Gaps   3;

Qy   254 GTTTTGCGAGCAGCTGGGACATTGTCAGCAATGCTGAAAAGCGAAAGGAGTGATGAGATG 313
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GTTTTGCGAGCAGCTGGGACATTGTCAGCAATGCTGAAAAGCGAAAGGAGTGATGAGATG 60

Qy   314 AAACGAATGGCAGAGAAGAATGAGCTGAGCGGTCAGTAAATGAGTTTCTGTCCAAGCTGCAA 373
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 AAACGAATGGCAGAGAAGAATGAGCTGAGCGGTCAGTAAATGAGTTTCTGTCCAAGCTGCAA 120

Qy   374 GATGACCTCAAGAGGAGGCAATGAATACTATGATGTGTAGCCGATGCCAAGGAAAGCATAGG 433
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GATGACCTCAAGAGGAGGCAATGAATACTATGATGTGTAGCCGATGCCAAGGAAAGCATAGG 180

Qy   434 AGGTTTGAAATGGACCGGAACCTAAGAGTGCCAGATACTGTCTGAGTGTAATAGGCTG 493
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 AGGTTTGAAATGGACCGGAACCTAAGAGTGCCAGATACTGTCTGAGTGTAATAGGCTG 240

Qy   494 CATCCTGCTGAGGAAGGAGACTTTTGGGCAGCATGTCAAGCACATGTGGCCTCAAGATCACC 553
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 CATCCTGCTGAGGAAGGAGACTTTTGGGCAGCATGTCAAGCACATGTGGCCTCAAGATCACC 300

Qy   554 TACTTTGCACTGATGGATGGAAAGGTGTATGACATCACAGAGTGGGCTGGATGCGCCAGCGT 613
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 TACTTTGCACTGATGGATGGAAAGGTGTATGACATCACAGAGTGGGCTGGATGCGCCAGCGT 360

Qy   614 GTAGGTATCTCCCCAGATACCCCAGAGTCCCTATCACATCTCATTTGGTTCTCGGATT 673
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 GTAGGTATCTCCCCAGATACCCCAGAGTCCCTATCACATCTCATTTGGTTCTCGGATT 420

Qy   674 CCAGGCACCCAGAGGCGGCGGCAGAGAGCCACCCCAGATGCCCCTCCTGCTGATCTTCAGGAT 733
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 CCAGGCACCCAGAGGCGGCGGCAGAGAGCCACCCCAGATGCCCCTCCTGCTGATCTTCAGGAT 480

Qy   734 TTCTTGAGTCGGATCTTTCAAGTGCCAGATGCCCAATGGGCCAGATGCCAATGGGCCAGATTCTTTGCA 793
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 TTCTTGAGTCGGATCTTTCAAGTGCCAGATGCCAATGGGCCAGATGCCAATGGGCCAGATTCTTTGCA 540

Qy   794 GCTCCTCAGCCTGCCCTGGAGCCGCTGCAGCCTCTAAGCCCAACAGCACAGTACCCAAG 853
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 GCTCCTCAGCCTGCCCTGGAGCCGCTGCAGCCTCTAAGCCCAACAGCACAGTACCCAAG 600

Qy   854 GGAGAAGCCAAACCTAAGCGGCGGGAAGAAAGTGAGGAGG-CCCTTCCAACGTTGATGCCC 912
          |||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db   601 GGAGAAGTCAAACCTAAGCGGCGGGAACAAAGTGAGGAGGCCCTTCCAACGTTGATGCCC 660

Qy   913 CTTCTCTTTCCTCAAATCAATGTAGCA-CAGGATGGAGTT 971
          ||||||||||||||||||| || ||||||||||||||||
Db   661 CTTCTCTTTCCTCCAATCCATGTCAGGGGAGTTAACACCAGGATGGAGTT 720

Qy   972 T-GATTTATCCCTCCCCCAACACCTAGGAACTGAATCTTTTTCTTTTATTTTTGA 1030
          | |||||||||||||||| || |||||||||||||||| |||| |
Db   721 TGGATTTATCCCTCCCCCTCCCCCCAACCTAAGGAAAACTGAATCCTTCCCTTTTCA 780

RESULT 14
BM452174
LOCUS   BM452174   995 bp   mRNA   linear   EST 05-FEB-2002
DEFINITION AGENCOURT_6386043 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:5526530
            5', mRNA sequence.
ACCESSION   BM452174
VERSION   BM452174.1 GI:18501214
KEYWORDS   EST.
SOURCE   Homo sapiens (human)

```
Qy   551 ACCTACTTTGCACTGATGATGGAAAGGTGTATGACATCACAGAGTGGGCTGGATGCCAG  610
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 ACCTACTTTGCACTGATGATGGAAAGGTGTATGACATCACAGAGTGGGCTGGATGCCAG  240

Qy   611 CGTGTAGTATCTCCCAGATACCCACAGAGTCCCTATCACATTCATTTGGTTCTCGG  670
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 CGTGTAGGTATCTCCCAGATACCCACAGAGTCCCTATCACATTCATTTGGTTCTCGG  300

Qy   671 ATTCCAGGCACCAGAGGGCGGCAGAGAGCCACCCCAGATGCCCCTCCTGCTGATCTTCAG  730
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 ATTCCAGGCACCAGAGGGCGGCAGAGAGCCACCCCAGATGCCCCTCCTGCTGATCTTCAG  360

Qy   731 GATTTCTTGAGTCGGATCTTTCAAGTACCCCCAGGCCCAATGGGCCCAATGGGAACTTCTTT  790
         |||||||||||||||||||||||||||||||| |||| |||||||||||||| ||||||
Db   361 GATTTCTTGAGTCGGATCTTTCAAGTACCCCCA-GGCAGATGCCCAATGGGAACTTTCTTT  419

Qy   791 GCAGCTCCTCAGCCTGCCCCTGGAGCCCGCTGCAGCCTCTAAGCCAACAGCACAGTACCC  850
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   420 GCAGCTCCTCAGCCTGCCCCTGGAGCCCGCTGCAGCCTCTAAGCCAACAGCACAGTACCC  479

Qy   851 AAGGGAGAAGCCAAACCTAAGCGGCGAAGAAAGTGAGGAGGCCCTTCCAACGTTGATGC  910
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   480 AAGGGAGAAGCCAAACCTAAGCGGCGAAGAAAGTGAGGAGGCCCTTCCAACGTTGATGC  539

Qy   911 CCCTTCCTTCTTCCTCAAATC-AATGTCAGGGAGTCAAAAGGGCTGTAGCACAGGATGGAG  969
         |||||||||||||||||||||  |||||||||||||||||||||||||||||||||||||
Db   540 CCCTTCCTTCTTCCTCAAATCAAATGTCAGGGAGTCAAAAGGGCTGTAGCACAGGATGGAG  599

Qy   970 TTTGATTTATCCTCCTCCCCCACACCTAGGAACTGAATCTTTTTCTTTTATTTTTG  1029
         ||||||||||||||||||||   |||||||||||||||||||   ||||||||||||
Db   600 -TTGATTTATCCTCCTCCCCCAA-AACTAGGAACTGAATCTTT---TCTTTTATTTTTG  654

Qy  1030 AGATGGAGTCTTGCTCTGTTGCCCAGTGGAGTGCAGTGGTGATCTCAGCTTACTGCA  1089
         |||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db   655 AGATGGAGTCTTGCTCTGTTGCCCAGTGGAGTGCAGTGGTGATCTCAGCTTAATGGAA  714

Qy  1090 ACCTCTGTC  1098
         | ||||||
Db   715 CTCTGTTTC  723
```

Search completed: December 1, 2003, 09:36:56
Job time : 2789 secs

```
Qy  1011 TTTTTCTTTTTTATTTTTGAGATGAGTCTTGCTCTGTTGCCCAGTGGAGTGCAGTGGT  1070
         ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db   541 TTTTTCTTTTTTATTTTTGAGATGGAGTCTTGCTCTGTTGCCCAGTGGAGTGCAGTGGT  600

Qy  1071 GTGATCTCAGCTTACTGCAACCTCCGTCTCCCGGGTTCAAGCAATTCCCCATCTCCAGCC  1130
         |||||||||||||||||||||||| |||||||||| |||| ||||||||||||||||||
Db   601 GTGATCTCAGCTTACTGCAACCTCTGTCTCCCGGGGTTCAGCAATTCCCCATCTCCAGCC  660

Qy  1131 TCCTGAGTAGCTGGGAT  1147
         |||||||||||||||||
Db   661 TCCTGAGTAGCTGGGAT  677
```

RESULT 15
BE619407
LOCUS   BE619407   723 bp   mRNA   linear   EST 20-OCT-2000
DEFINITION 601473229F1 NIH_MGC_68 Homo sapiens cDNA clone IMAGE:3876199 5',
            mRNA sequence.
ACCESSION   BE619407
VERSION   BE619407.1 GI:9890345
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 723)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP/Gazdar
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9636 row: j column: 08
            High quality sequence stop: 707.
FEATURES   Location/Qualifiers
     source   1..723
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3876199"
                     /tissue_type="large cell carcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_68"
                     /note="Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.8 kb. Library constructed by Life
                     Technologies."
BASE COUNT   181 a   181 c   189 g   172 t
ORIGIN

Query Match   56.2%; Score 644.2; DB 10; Length 723;
Best Local Similarity 97.3%; Pred. No. 1.6e-22;
Matches 709; Conservative   0; Mismatches   13; Indels   7; Gaps   5;
```

```
Qy   371 CAAGATGACCTCAAGGAGGCAATGAATGAATCTATGATGTGTAGCCGATGCCAAGGAAAGCAT  430
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 CAAGATGACCTCAAGGAGGCAATGAATGAATCTATGATGTGTAGCCGATGCCAAGGAAAGCAT   60

Qy   431 AGGAGGTTTGAAATGGACCGGAACCTAGGAGTGCCAGATACTGCTGCTGAGTGTAATAGG  490
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 AGGAGGTTTGAAATGGACCGGAACCTAGGAGTGCCAGATACTGCTGCTGAGTGTAATAGG  120

Qy   491 CTGCATCCTGCTGAGGAAGGAGACTTTGGGCAGAGTCAAGCATGTTGGGCCTCAAGATC  550
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 CTGCATCCTGCTGAGGAAGGAGACTTTGGGCAGAGTCAAGCATGTTGGGCCTCAAGATC  180
```

```
Qy   541  CCTCAAGAGATCACCTACTTTGCACTGATGATGGAAAGGTGTATGACATCACAGAGTGGGC  600
          |||||||||| ||| ||||||||||||| ||-|| ||||||||| |||| |||||||||||
Db 110652  CCTCAAGAGATCTCCTGTTTTGCACTGACTGACTGAACGGTGTATGTCATCACAGAGTGGCC 110711

Qy   601  TGGATGCCAGCGTGTAGGTATCTCCCCAGATACCCACAGAGTCCCTATCACATCTCATT  660
          ||||||| |||                       ||||||| ||-|-|| |||||||||||
Db 110712  TGGATGTCAG-----------TCCAGATACTCACAGGATTCCTTACCACTTCTCATT 110757

Qy   661  TGGGTCTCGGATTCCAGGCACCAGAGGGCGGCAGAGAGCCAGATGCCCCTCCTGC  720
          ||||| |||||||||||| ||||-| ||||| |||| |||| ||-|||||| ||||
Db 110758  TGTTTCTTGGGTACTGGGCACCAGTGATTGGCCTAGGGTCACTCCAGAGTCTTCTCCTGC 110817

Qy   721  TGATCTTCAGGATTTCTTGAGTCGGATCTTTCAAGTACCCCCAGGGCAGATGCCCAATGG  780
          ||| || ||| |||||||| |||-| || |||| ||||| |||| |||| ||| |||
Db 110818  TGAGCTGCAGGATTTCTTGAGCCAGATCTTTCAAGTACCTCCCGGGCCGATGTCCAGTGG 110877

Qy   781  GAACTTCTTTGCAGCTCCTCAGCCTGCCCCTGGAGCCGCTGCAGCCTCTAAGCCCAACAG  840
          ||-|||||||||||||| |||||||||| ||||||-| || |||||||||||||| ||||
Db 110878  GACCTTCTTTGCCTCACCTCACCCTGGCCCCTGGAACCACTTCAACTTCAAGGCTCAACAG 110937

Qy   841  CACAGTACCCAGGGAGAGCCAAACCTAAGCGCGGAGAAAGTGAGGAGGCCCTTCCA  900
          |||||||||||||||||||||| ||-||| ||| ||| |||||| |||| |||||||
Db 110938  TTCAGTACCCAAGAAGAGCCAAACATAAACGGTAGAGAAGTGAGGTGGCCCCTCCTA 110997

Qy   901  ACGTTGATGCCCCTT---CTCTTTCCTCAAATCAATGTCAGGGAGTCAAAAGGGCTGT-  955
          ||| ||| ||||||||| |||||||| |||||| ||||| ||||||| ||||||||||||
Db 110998  ACGATGACACCCCCTTCTCCTCCCTTCCTCAAATCAATGTCAGGGAGTCAAAAGGGCTGTGT 111057

Qy   956  --AGCACACGATGGAGTTTGATTTAT  979
            |||| ||||| || |||||||||||
Db 111058  ACAGCACGGGATAGAATTTGATTTAT 111083
```

Search completed: December 1, 2003, 08:50:25
Job time : 4382 secs

Lorensuhewa,L., Louiseged,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Margum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwaokelemeh,O., Okwuonu,G., Olarnpunsagon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rockey,T., Rojas,A., Rose,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
  TITLE   Direct Submission
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 247309)
  AUTHORS   Worley,K.C.
  TITLE   Direct Submission
  JOURNAL   Submitted (23-OCT-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3 (bases 1 to 247309)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE   Direct Submission
  JOURNAL   Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT   On May 10, 2003 this sequence version replaced gi:25012326.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------- Genome Center
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
            -------- Project Information
            Center project name: GFIX
            Center clone name: CH230-57L7
            -------- Summary Statistics
            Assembly program: Atlas 3.0;
            Consensus quality: 233539 bases at least Q40
            Consensus quality: 235537 bases at least Q30
            Consensus quality: 236797 bases at least Q20
            Estimated insert size: 246711; sum-of-contigs estimation
            Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            ----
            * NOTE: Estimated insert size may differ from sequence length
              (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is

            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *   1   246159: contig of 246159 bp in length
            *   246160   246259: gap of unknown length
            *   246260   247309: contig of 1050 bp in length.
FEATURES   Location/Qualifiers
     source   1..247309
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-57L7"
     misc_feature   1..1645
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature   6457..7267
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH299097"
     misc_feature   239013..239869
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH299095"
     misc_feature   243192..246159
                     /note="wgs_end_extension
                     clone_end:T7"
BASE COUNT   69270 a 46265 c 47122 g 74568 t 10084 others
ORIGIN

Query Match   57.6%; Score 661.2; DB 2; Length 247309;
Best Local Similarity 81.8%; Pred. No. 3.3e-202;
Matches 807; Conservative 0; Mismatches 158; Indels 21; Gaps 3;

Qy   1 GCCTTGGGTCAAGCAGAATATTAATAGGGAGGGAATGCCACCTGTACTGGTAGTGGGGCTA 60
            |||| ||||||||||||| ||| || |||| ||| ||||| ||| ||||||||| ||| ||
Db   110112 GCTCTGAGTCATGCAGAGACTAAGAAACAGAGCAATGCCCCTGTACTGGGTCTAGCGGGAGATA 110171

Qy   61 CTGCCAGCTGAAGAGGAAGTGGCTCGACTCTTGACACCATGGCTGGGGTTCCTGAGGATGA 120
            ||||||||||||||||||||| || |||||||||| | ||||||||| ||||||||||||
Db   110172 CTGCCAGCTGAAGAGGAAGTAGCCCGACTCTTGATTATGGCTGGGTTTCCTGAGGATGA 110231

Qy   121 GCTAAACCCTTTCCATGTACTGGGGTTGAGGCCACAGCATCAGATGTTGAACTGAAGAA 180
            |||||||||||||||||| || ||||||||| | ||||| || |||| |||||||||||
Db   110232 ACTAAACCCTTTCCATGTGCTAGGGGTTGAAGCTACAGCGTCCGATATTGAACTGAAGAA 110291

Qy   181 GGCCTATAGACAGCTGGCAGTGATGGTTCATCCTGACAAAAATCATCACCCCGGGCTGA 240
            |||||||||||| || || | || ||| ||||||| ||||||||| ||| ||||||||
Db   110292 GGCCTATAGACAACTAGTGTAATGATCCATCCTGATAAAAATCACAATCCCAAGGCTGA 110351

Qy   241 GGAGGCCTTCAAGGTTTGGGAGCAGCTGGGACATTGTCAGCAATGCTGAAAAGCGAAA 300
            |||||||||||| |||||| | ||||| |||||||||||||||| | |||||||| ||
Db   110352 GGAGGCCTTCAAAGTTTGCGGGCAGCTTGGGACATTGTCAGCAACCCAGAGAGGCAGAA 110411

Qy   301 GGAGTATGAGATGAAACGAATGCAGAGAATGAGCTGAGCCGGTCAGTAAATGAGTTTCT 360
            ||| ||||||||| || |||||||||||| |||||| ||||||| ||| |||||||||||
Db   110412 GGAATATGAGATAAATGAATGCAGAGAGTGAGCTCAGCCGGTTAGTGAATGAGTTTCT 110471

Qy   361 GTCCAAGCTGCCAAGATGACCTCAAGGAGGCAATGAATGATGTGTAGCCGATGCCA 420
            ||| ||||| |||||||||||||||| ||| | |||| || || |||||||||||
Db   110472 GTCGAAGCTACAAGATGACCTCAAGAAGGACTGGGAACCCAAGAGTGCCAGATACTGTGCTGA 110531

Qy   421 AGGAAAGCATAGGAGGTTTGAAATGACGGGAACCTAAGAGTGCCAGATACTGCTGA 480
            ||||| ||||| ||||||||||| || |||| ||| || || ||||| ||||||||
Db   110532 AGGAAAGAATAGGGGGTTTGAAAAGGACTGGGAACCCAAGAGTGCCAGATACTGCTGA 110591

Qy   481 GTGTAATAGGCTGCATCCTGCTGAGGAACGAGGACTTTTGGCAGAAGGAGACATGTTGGG 540
            |||| |||||| | ||||||||| |||||||||| |||||| || || |||||||||||
Db   110592 CTGTAACAGGCTTCATCCTGCTGAGGAACGAGGAAGAGTCTTCTGGGTAGGAGTCAAGGATGTTGGG 110651

Qy   901 ACGTTGATGCCCCTTCTCTT--TCCTCAAATCAATGTCAGGGAGTCAAAAGGGCTGT--- 955
          ||||||| ||| |||||||| |||||| |||||| ||||||||||||||||||||
Db 142449 ACGATGAAACCACCTTCTTCTCTTCCTTCTTCTTCAGTCAATATCAGGGAGTCAAAAGGGCTGTGTA 142390

Qy   956 -AGCACAGGATGGAGTTTGATTTAT 979
           |||||||||||||||||||||||
Db 142389 CAGCACAGGATGGAGTTTGATTTAT 142365

RESULT 14
AF141342
LOCUS   AF141342   754 bp   mRNA   linear   PRI 14-JAN-2001
DEFINITION Homo sapiens LYST-interacting protein LIP6 mRNA, complete cds.
ACCESSION   AF141342
VERSION   AF141342.1 GI:12239359
KEYWORDS   .
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 754)
  AUTHORS   Tcherney,V.T., Mansfield,T.A., Giot,L., Kumar,A.M., Nandabalan,K.,
            Li,Y., Mishra,V.S., Detter,J.C., Rothberg,J.M., Wallace,M.R.,
            Southwick,F.S. and Kingsmore,S.F.
  TITLE   Interactions of the Chediak-Higashi lysosomal-trafficking regulator
            protein with SNARE complex and signal transduction proteins
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 754)
  AUTHORS   Tcherney,V.T., Mansfield,T.A., Giot,L., Kumar,A.M., Nandabalan,K.,
            Li,Y., Mishra,V.S., Detter,J.C., Rothberg,J.M., Wallace,M.R.,
            Southwick,F.S. and Kingsmore,S.F.
  TITLE   Direct Submission
  JOURNAL   Submitted (08-APR-1999) Full-Length Cloning, CuraGen Corporation,
            12085 Research Drive, BDI, Alachua, FL 32615, USA
FEATURES   Location/Qualifiers
     source   1..754
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
     CDS   11..586
                     /codon_start=1
                     /product="LYST-interacting protein LIP6"
                     /protein_id="AAG49445.1"
                     /db_xref="GI:12239360"
                     /translation="MKRMAENELSRSVNEFLSKLQDDLKEAMNTMMCSRCCQGKHRRFE
                     MDREPKSARYCAECNRLHPAEEGDFWAESSMGLKITYFALMDGKVYDITEWAGCQRV
                     GISPDTHRVPVHISFGSRIPGTRGRQRATPDAPPADIQDFLSRIFQVPPGQMPMGTSL
                     QLLSLPEPLQPLSPTAQYPREKPNLSGGRK"
BASE COUNT   193 a   185 c   200 g   176 t
ORIGIN

Query Match   64.8%; Score 743; DB 9; Length 754;
Best Local Similarity   99.9%; Pred. No. 5.6e-229;
Matches 754; Conservative   0; Mismatches   0; Indels   0; Gaps   1;

Qy   301 GGAGTATGAGATGAAACAATGCCAGAGAATGAGCTGAGCCGGTCAGTAAATGAGTTTCT 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GGAGTATGAGATGAAACAATGCCAGAGAATGGCTGAGCTGAGCCGGTCAGTAAATGAGTTTCT 60

Qy   361 GTCCAAGCTGCAAGATGACCTCAAGGAGGCAATGAATACTATGATGTGTAGCCGATGCCA 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GTCCAAGCTGCAAGATGACCTCAAGGAGGCAATGAATACTATGATGTGTAGCCGATGCCA 120

Qy   421 AGGAAAGCATAGGAGGTTTGAAATGCCAGGGAACCTAAGAGTGCCAGACTACTGTGCTGA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 AGGAAAGCATAGGAGGTTTGAAATGGCAGGGAACCTAAGAGTGCCAGACTACTGTGCTGA 180

Qy   481 GTGTAATAGGCTGCATCCTGCTGAGGAAGGAGACTTTTGGGCAGAGTCAAGCATGTGGG 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 GTGTAATAGGCTGCATCCTGCTGAGGAAGGAGACTTTTGGGCAGAGTCAAGCATGTGGG 240

Qy   541 CCTCAAGATCACCTACTTTGCACTGATGGATGGAAAGGTGTATGACATCACAGAGTGGGC 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 CCTCAAGATCACCTACTTTGCACTGATGGATGGAAAGGTGTATGACATCACAGAGTGGGC 300

Db   241 CCTCAAGATCACCTACTTTGCACTGATGGATGGAAAGGTGTATGACATCACAGAGTGGGC 300

Qy   601 TGGATGCCAGCGTGTAGGTATCTCCCAGATACCACCACAGAGTCCCTATCACATCTCATT 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 TGGATGCCAGCGTGTAGGTATCTCCCAGATACCACCACAGAGTCCCTATCACATCTCATT 360

Qy   661 TGGTTCTCGGATTCCAGGCACCAGAGGGCGGCGAGAGAGCCACCCCAGGGCCCTCCTGC 720
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 TGGTTCTCGGATTCCAGGCACCAGAGGGCGGCGAGAGAGCCACCCCAGGGCCCTCCTGC 420

Qy   721 TGATCTTCAGGATTTCTTGAGTCGGATCTTTCAAGTACCCCCAGGGCCAGATGCCAATGG 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 TGATCTTCAGGATTTCTGAGTCGGATCTTTCAAGTACCCCCAGGGCCAGATG-CCAATGG 479

Qy   781 GAACTTCTTGCAGCTCTTCAGCCTGCCCTGCGCCGCTCAGCCTTAAGCCAACAG 840
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   480 GAACTTCTTGCAGCTCTTCAGCCTGCCCTGCGCCGCTCAGCCTTAAGCCAACAG 539

Qy   841 CACAGTACCCAAGGAGGAGCAAACCTAAGCGGCCGGAAGAAAGTGAGGAGGCCCTTCCA 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   540 CACAGTACCCAAGGAGGAGCAAACCTAAGCGGCCGGAAGAAAGTGAGGAGGCCCTTCCA 599

Qy   901 ACGTTGATGCCCCTTCTCTTCCTCTTCCTCCCCCAACAATCAATGTCAGGGAGTCAAAAGGGCTGTAGAC 960
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   600 ACGTTGATGCCCCTTCTCTTCCTCTTCCTCCCCCAACAATCAATGTCAGGGAGTCAAAAGGGCTGTAGCAC 659

Qy   961 AGGATGGAGTTTGATTTATCCTCCCCCAACAACTAGAACTGAATCTTTTTCTTTT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   660 AGGATGGAGTTTGATTTATCCTCCCCCAACAACTAGGAACTGAATCTTTTTCTTTT 719

Qy   1021 TATTTTTTGAGATGGAGTCTTGCTCTGTTGCCCAG 1055
          |||||||||||||||||||||||||||||||||||
Db   720 TATTTTTTGAGATGGAGTCTTGCTCTGTTGCCCAG 754

RESULT 15
AC097837
LOCUS   AC097837   247309 bp   DNA   linear   HTG 10-MAY-2003
DEFINITION Rattus norvegicus clone CH230-57L7, WORKING DRAFT SEQUENCE, 2
            unordered pieces.
ACCESSION   AC097837
VERSION   AC097837.7 GI:30520389
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE   Rattus norvegicus (Norway rat)
  ORGANISM Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 247309)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,D., Chen,G., Chen,R., Chen,Y., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Forbes,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,K., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,

```
            TITLE      Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            JOURNAL    Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            REFERENCE  Weinstock,G. and Gibbs,R.
            AUTHORS    Direct Submission
                       Unpublished
            TITLE      2 (bases 1 to 263371)
            JOURNAL    Worley,K.C.
                       Direct Submission
                       Submitted (23-OCT-2001) Human Genome Sequencing Center, Department
                       of Molecular and Human Genetics, Baylor College of Medicine, One
                       Baylor Plaza, Houston, TX 77030, USA
            REFERENCE  3 (bases 1 to 263371)
            AUTHORS    Worley,K.C.
            TITLE      Direct Submission
            JOURNAL    Submitted (10-OCT-2002) Human Genome Sequencing Center, Department
                       of Molecular and Human Genetics, Baylor College of Medicine, One
                       Baylor Plaza, Houston, TX 77030, USA
            COMMENT    On Oct 10, 2002 this sequence version replaced gi:21953452.
                       The sequence in this assembly is a combination of BAC based reads
                       and whole genome shotgun sequencing reads assembled using Atlas
                       (http://www.hgsc.bcm.tmc.edu/projects/rat/).Each contig described
                       in the feature table below represents a scaffold in the Atlas
                       assembly (a 'contig-scaffold'). Within each contig-scaffold,
                       individual sequence contigs are ordered and oriented, and separated
                       by sized gaps filled with Ns to the estimated size. The sequence
                       may extend beyond the ends of the clone and there may be sequence
                       contigs within a contig-scaffold that consist entirely of whole
                       genome shotgun sequence reads. Both end sequences and whole genome
                       shotgun sequence only contigs will be indicated in the feature
                       table.
                       ------- Genome Center
                               Center: Baylor College of Medicine
                               Center code: BCM
                               Web site: http://www.hgsc.bcm.tmc.edu/
                               Contact: hgsc-help@bcm.tmc.edu
                       ------- Project Information
                               Center project name: TUUD
                               Center clone name: CH230-2H10
                       ------- Summary Statistics
                               Assembly program: Phrap; version 0.990329
                               Consensus quality: 243674 bases at least Q40
                               Consensus quality: 246830 bases at least Q30
                               Consensus quality: 248548 bases at least Q20
                               Estimated insert size: 247731; sum-of-contigs estimation
                               Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
                       -------
                       * NOTE: Estimated insert size may differ from sequence length
                       *       (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
                       * NOTE: This is a 'working draft' sequence. It currently
                       *       consists of 3 contigs. The true order of the pieces
                       *       is not known and their order in this sequence record is
                       *       arbitrary. Gaps between the contigs are represented as
                       *       runs of N, but the exact sizes of the gaps are unknown.
                       *       This record will be updated with the finished sequence
                       *       as soon as it is available and the accession number will
                       *       be preserved.
                       *           1     258056: contig of 258056 bp in length
                       *       258057     258156: gap of unknown length
                       *       258157     261384: contig of 3228 bp in length
                       *       261385     261484: gap of unknown length
                       *       261485     263371: contig of 1887 bp in length.
            FEATURES             Location/Qualifiers
            source            1..263371
                              /organism="Rattus norvegicus"
                              /mol_type="genomic DNA"
                              /db_xref="taxon:10116"
                              /clone="CH230-2H10"
            misc_feature      1..1125
                              /note="wgs_contig"
            misc_feature      2121..4146
                              /note="wgs_contig"
            misc_feature      5480..6935
                              /note="wgs_contig"
```

```
            misc_feature      256941..258056
                              /note="wgs_contig"
            misc_feature      258157..259735
                              /note="wgs_contig"
            misc_feature      259889..261384
                              /note="wgs_contig"

            BASE COUNT     75390 a   46168 c   46816 g   81072 t   13925 others
            ORIGIN

            Query Match                          66.8%;  Score 765.8;  DB 2;  Length 263371;
            Best Local Similarity   87.5%;  Pred. No. 4e-236;
            Matches  862;  Conservative   0;  Mismatches 117;  Indels   6;  Gaps   2;

Qy       1 GCCTTGGTCAAGCAGAATATTAATAGGCAGGGAATGCACCTGTAGCTAGTGGGCGCTA    60
           ||||||||||||||||| |||| || || ||||||||||||| |||| ||| || |||
Db  143349 GCCCTGGGTCAAGCAGAGACCAGAGAGGACTCAGACAGGGCACTGCCCATGTAAATAGCGGGCGATA 143290

Qy      61 CTGCCAGCCTGAAGAGGAGAAGTGGCTCGACTCTTGACCATGGCTGGGGTTCCTGAGGATGA   120
           ||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db  143289 CTGCCAGCCTGAAGAGGAGAAGTCGCCCGACTCTTGACCATGGCTGGGGTTCCTGAAGATGA 143230

Qy     121 GCTAAACCCTTCCATGTACTGGGGGTTGAGCGCCAGCATCAGATGTTGAACTGAGAA   180
           ||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db  143229 ACTAAACCCTTTCATGTGCTGGGGGTTGAAGCTACAGCATCCGATATTGAACTAAAGAA 143170

Qy     181 GGCCTATAGACAGCTGGCAGTGATGGTTCATCCTGACAAAAATCATCATCCCGGGCTGA   240
           |||||||||||| |||||||||||||||||||||||||||   || ||||||||||||||
Db  143169 AGCCTATAGGCAGCTAGCGGTGATGGTTGATGGTCCATCCTGATAAAAATCACCATCCCCGGGCTGA 143110

Qy     241 GGAGGCCTTCAAGGTTTTGCGGACGCAGCTGGGACATTGTCAGCAATGCTGAAAAGCGAAA   300
           ||||||||||||||||||||||||||||||||||||||||||||| || ||| || | ||||
Db  143109 GGAGGCCTTTAAAGTTTTGCGGACGCAGCTGGGACATTGTCAGCAATCCAGAGAGCGCGAA 143050

Qy     301 GGAGTATGAGATGAAACGATGCAGAGAATGAGTCGAGCCGGTCAGTAAATGAGTTTCT   360
           ||| ||||||| ||||||||||||||||| ||||||||||| || ||||| ||||||||
Db  143049 GGAATATGAGATGAAACGGATGCAGAGAGAATGAGCTCAGCCGGCTCGGTGAATGAGTTTCT 142990

Qy     361 GTCCAAGCTGCAAGATGACCTCAAGGAGGCAATGGAATACTATGATGTGTAGCCGATGCCA   420
           |||||||||||||||||||||||||||| || ||||| || || || ||||||||||||||
Db  142989 TTCGAAGCTACAAGATGATTTCAAGGAGCCAATGAATAACGATGATGATGTGCAGCCGGTGCCA 142930

Qy     421 AGGAAAGCATAGGAGGTTGAAATGACCGGGAACCTAAGAGTGCCAGATACTGTCGTGA   480
           |||||| |||||||| |||||||||||||||||||||||||||||||||||||||||||
Db  142929 AGGGAAGCATAGGAGGTTGAAATGACCGGGAACCCAAGAGTGCCAGATACTGTCGTGA 142870

Qy     481 GTGTAATAGGCTGCATCCTGCTGAGGAAGGAGACTTTTGGCAGCAGAGTCAAGCATGTTGGG   540
           ||||| ||| || ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  142869 GTGTAACAGGTGCATCCTGCTGCTGAGGAAGGAGAGACTTCTCGGGCAGAGTCAAGCATGTTGGG 142810

Qy     541 CCTCAAGATCACCTACTTTGCACTGATGAAGGTGTATGACATCACAGAGTGGGC   600
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  142809 CCTCAAGATCACCTACTTTGCACTGATGGGAAGGTGTATGACATCACAGAGTGGTC 142750

Qy     601 TGGATGCCAGCGTGTAGGTATCTCCCAGATACCCACAGAGTCCCTATCACATCTCATT   660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db  142749 TGGATGCCAGCGTGTGGGCATCTCCCAGATACCCACAGAGTCCCTATCCTTACCACACTCATT 142690

Qy     661 TGGTTCTCGGATTCCAGGCACCACCAGAGGGCCGCAGAGGCCACCCCAGATGCCCTTCCTGC   720
           ||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db  142689 TGGTTCTCGGTTACCAGGCACCACCAGAGCGGCGGCAGACAACAGCCACTCCAGATCCCCTTCCTGC 142630

Qy     721 TGATCTTCAGGATTTCTTGAGTCGATCTTTCAAGTACCCCAGGCAGATGCCCAATGG   780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  142629 TGACCTGCAGGATTTCTTGAGCCGGATCTTTCAAGTACCTTCAAGTACCCAGGCCGATGGTCGATGGTC 142570

Qy     781 GAACTTCTTTGCAGCTCCTCAGCCTGCGCCCTGAGAGCCCGCTCAAGCCCAACAG   840
           ||||||||||||||||||||||||||||||||||||||||| | | ||||||||||
Db  142569 GAACTTCTTTGCCGACTCACCTCACCCTGGGACTACTTCAACCTCCAGGCCCAACAG 142510

Qy     841 CACAGTACCCAAGGAGAGCCAAACCTAAGCGGCGGAAGAAAGTGAGGAGGCCCTTCCA   900
           ||||||| | |||||| || ||| |||| || ||  ||||||||||||| ||||||||
Db  142509 TTCAGTATCCAAGGAGAGCCAAACCTAAACGTTGGAGAAGAAGTGAGTGGGCCCTTCCA 142450
```

```
          /gene="Drip78"
          /note="DRIP78; ER-membrane protein; zinc finger motif"
          /codon_start=1
          /product="dopamine receptor interacting protein"
          /protein_id="AAK56240.1"
          /db_xref="GI:14194055"
          /translation="MAQKHPGERGLCGVHHSGGSSLLITSGSSVDPEILSFSGLRDSKE
          TAPNGTRCLKEHSDPKCTQPSNPAHWSDPSHGPPRGPGPPREGYPDESETCSSESGV
          DQELSRENTGYQEDGSPSFLPIPSACNCQGSPGVPEGTCSEEGDGSSSSFCHHCTSP
          ALGEDEELEEYDDEEPLKFPSDFSRVSSGKKPPSRRQRHRFLTKEDVRDSGRRDPKA
          PGRHRLAEVRQITPEGLGWGVEGLCQLGQAGFWLIELLVLVGEYVETFCGYLIYACR
          KLKGSDLDLFRIWGVWARRLGGWARVMVMFQFLSQSFFSVAGLFIRLLRVVGAFLLAL
          ALFLGCLGLGWRFLVGLGDRLGWRGKAAWLFSWLDSPALHHFLTLLKDSRPWQQLVRV
          IQWGWLELPWVKQRTQRQGTANVPSGRYCQPEEEVARLLTMAGVPEDELNPFHVLGVE
          ATASDIELKKAYRQLAVMVHPDKNHHPRAEEAFKVLRAAWDIVSNPERRKEYEMKRMA
          ENELSRSVNEFLSKLQODLKEAMNTMMCSRCQGKHRRFEMDREPKSARYCAECNRLHP
          AEBGDFWAESSMLGLKITYFALMDGKVTDITEWAGCQRVGISPDTHRVPYHISFGSRV
          PGTSGRQRATPESPPADLQDFLSRIFQVPPGPMSNGNFFAAPHPGGTTSTSRPNSSV
          PKGEAKPKRKKKVRRPFQR"

BASE COUNT    565 a    582 c    723 g    522 t
ORIGIN
```

```
Query Match       68.2%;  Score 781.8;  DB 10;  Length 2392;
Best Local Similarity 88.5%;  Pred. No. 1.7e-241;
Matches  872;  Conservative   0;  Mismatches 107;  Indels   6;  Gaps    2;

Qy      1 GCCTGGGTCAAGCAGCAGAATATTAATAGGCAGGGGAATGCACCTGTAGCCACTGTGGCCGCTA     60
            ||| ||||||||||||||||  ||  ||||| ||||| | |||||| |||| ||||||||||
Db   1360 GCCCTGGGTCAAGCAGCAGAGGACTCAGAGGCAGGGCACTCAGCAGTTCCTGAGGATGA   1419

Qy     61 CTGCCAGCCTGAAGAGGAGAGTGGCTCGACTTCTTGACCATGCTCGGGGTTCCTGAGGATGA    120
            ||||||||||||||||| |||||||||||||||||||||||| ||||||||||||||||||||
Db   1420 CTGCCAGCCTGAAGAGGAAGTCGCCCAGCTCTTGACCATGGCCAGGGGGTTCCTGAAGATGA   1479

Qy    121 GCTAAACCCTTTCATGTACTGGGGGTTGAGGCCACAGCATCAGATGTTGAACTGAAGAA    180
            | | ||||||||||||||||||||||||| ||||||||||||| ||||||||||||||
Db   1480 ACTAAAACCCTTTCATGTGCTGGGGGTTGAAGCTACAGCATCCGATATTGAACTAAAGAA   1539

Qy    181 GGCCTATAGACAGCTGGCAGTGGCTGATGGTTCATCCTGACAAAAATCATCATCCCCGGCTGA    240
            |||||||| |||||||||||||| ||||||||||||||||| |||||||||||||||||||||
Db   1540 GGCCTATAGGCAGCTAGCGGTGATGGTCCATCCTGATAAAAACCACCATCCCCGGGGTGA   1599

Qy    241 GGAGGCCTTCAAGGTTTGCGAGCAGCTTGGGACATTGTCAGCAATGCTGAAAAGCGAAA    300
            |||||||||| |||||||||||||||||||||||||| |||||||||||||||||||||
Db   1600 GGAGGCCCTTTAAAGTTTTGCGGGACCAGCTTGGGACCATTGTCAGCAATCCTGAGCAGGGGAA   1659

Qy    301 GGAGTATGAGATGAAAACGAATGCCAGAAGTAGCTGAGCCGGTCAGTAAATGAGTTTCT    360
            |||| ||||||||||||||| |||||||||||||||| | ||| |||||||||||||||
Db   1660 GGAATATGAGATGAAAACCGATGCCAGAGGATGAGCTCAGCCGCTCGGTGAATGAGTTTCT   1719

Qy    361 GTCCAAGCTGCAAGATGACCTCAAGGAGGCAATGAATACTATGATGAATACTACTGAGCCGATGCCA    420
            |||||||||||||| ||||||||||||||||||| ||||| |||||| |||||||||||||||
Db   1720 GTCGAAGCTACAAGATGACCTCAAGGAGGCAATGAATGAATACGATGATGTGCAGCCGTGCCA   1779

Qy    421 AGGAAAGCATAGGAGGTTTGAAATGGAACCTAAGAGGTCCAGATACTGTGCTGA    480
            ||| ||||||||||||||||||||||||||| |||| |||||||| ||||||||||
Db   1780 AGGGAAGCATAGGAGGTTTGAAATGGAACCGGGAACCCAAGAGTGCCAGATACTGTGCCGA   1839

Qy    481 GTGTAATAGGCTGCATCCTGCTGAGGAAGGAGACTTTTGGGCAGAGGAGTCAAGCATGTTGGG    540
            |||| ||||||||||||| ||||||||||||| |||||| ||||||||||||||||||||||||
Db   1840 GTGTAACAGGTTGCATCCTGCTGAGGAGGAGACTTCTGGGCAGAGGAGTCAAGCATGTTGGG   1899

Qy    541 CCTCAAGATCACCTACTTTGCACTGATGGATGGAAAGGTGTATGACATCAGAGAGTGGGC    600
            |||||||||||||||| ||||||||| |||||||||||||||||| |||||||||||||||
Db   1900 CCTCAAGATCACCTACTTTGCACTGATGGATGGAAAGGTGTATGACATCAGAGAGTGGGC   1959

Qy    601 TGGATGCCAGCGTGTAGGTATCTCCCCAGATACCCACCAGAGTCCCCTATCACATCTCATT    660
            |||||||||||||||||||| ||||| |||||||||||||| |||||||||||||||||||
Db   1960 TGGATGCCAGCGTGTGGGCATCTCCCCAGATACCCACCGAGTTCCTTACCACATCTCATT   2019

Qy    661 TGGTTCTCGGATTCCAGGCACCAGCGGCACCACCCCAGATGCCCTCCTGC    720
            |||||| ||| ||||||||||| ||| ||||||||||||| ||||||||||
Db   2020 TGGTTCGGGTACCAGCGCCGACAAGGGCCACTCCAGAGTCCCCTCCTGC   2079
```

```
Qy    721 TGATCTCAGATTTCTTGAGTCGGATCTTCAAGTACCCCAGGGCAGATGCCCAATGG    780
            |||||| |||||||||||||||||||||||||||||    |||| ||||||||||||
Db   2080 TGACCTGCAGGATTTCTTGAGCCGGATCTTTCAAGTACCTCCAGGGCCGATGTCCAATGG   2139

Qy    781 GAACTTCTTGCAGCCTCCTCAGCCTCCGCCCCTGGAGCCCTCTAAGCCAACAG    840
            |||||||||| |||||||||||| |||||| |||||||| |||||||||||||||
Db   2140 GAACTTCTTTGCCGCACCTCACCCTGCCCTGGGACTACTTCAACCTCCAGGCCCAACAG   2199

Qy    841 CACAGTACCCAAGGGAGAAGCCAAACCTAAGCGGCGGAGAGAAAGTGAGGAGGCCCTTCCA    900
            ||||||| |||||| ||| |||||||| | |||  |||||| |||||||||||| ||||| |
Db   2200 TTCAGTACCCAAGGGAGGAAGCCAAACCTAAACGGCGGAAGAAAGTGAGGCGCCCCTTCCA   2259

Qy    901 ACGTTGATGCGCCCTTCTT--TCCTCAAATCAATGTCAGGGAGTCAAAAGGGCTGT---    955
            ||  |||| | ||||| |||  |||||  |||||||| |||| |||||||||||||
Db   2260 AAGATGACACCACCCTTTCTTCTCTTCTCAGTCAATGTCAGTCAGGGAGTCAAAAGGGCTGTGTA   2319

Qy    956 -AGCACAGGATGGAGTTTGATTTAT    979
             |||||||||||| |||||||||||
Db   2320 CAGCGCAGGATGGAGTTTGATTTAT   2344
```

```
RESULT 13
AC098454/c
LOCUS       AC098454             263371 bp    DNA    linear    HTG 10-OCT-2002
DEFINITION  Rattus norvegicus clone CH230-2H10, WORKING DRAFT SEQUENCE, 3
            unordered pieces.
ACCESSION   AC098454
VERSION     AC098454.4  GI:23364557
KEYWORDS    HTG; HTGS_PHASE1; HTGS DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
ORGANISM    Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 263371)
AUTHORS     Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louiseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,Z., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newton,J., Newtson,A., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,V., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
```

This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis, GenomeScan gene prediction.

```
FEATURES             Location/Qualifiers
     source          1..2402
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="CZECH II"
                     /db_xref="taxon:10090"
                     /clone="MGC:19282 IMAGE:4016209"
                     /tissue_type="Mammary tumor metastatized to lung.
                     MMTV-LTR/Wnt1 model. Expression driven by an MMTV-LTR
                     enhancer."
                     /clone_lib="NCI_CGAP_Lu30"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     CDS             148..2259
                     /codon_start=1
                     /product="RIKEN cDNA 5730551F12 gene"
                     /protein_id="AAH11146.1"
                     /db_xref="GI:15029846"
                     /db_xref="LocusID:74330"
                     /translation="MAQXHPGERRLCGAHRSGGTSLSGTSGGSSVDPEILSFSGLRDSAE
                     TAPNGTRCLKEHSGPKYTQPPNPAHWSDPSHGPPRGPGPPRGGYPDESETGSEESGV
                     DQELSRENETGYOEDGSPSFLSIPSACNCQGSGVPEGTYSEEGDGSSSSLCHHCTSP
                     ALGEDEELEEEYDDEEPLKFPPDSFRSVSSGKKPLSRRQKHRFPIKEDVRDSGRREPKA
                     PGRHRLARKRSOTDKRRGLGLWGVEBLCQLGQAGFWWLIELLVIVGEYVETCGHLIYA
                     CRKLKGSDLDLFRVWVGVWARRIGGWARMFQFLSQSFFCVVGLLIRLRVVGAFLLL
                     ALALFLGCLQLGWRFSVGLGNRLGWRDKTAWLFSWLGSPALHHCLTLLKDSRPWQQLV
                     RLIQWGNGWQELPWKQRITKKQGNAPVASGRYCQPEEEVTRLLTMAGVPEDELNPFHVLG
                     VEATASDTELKKAYRQLAVMHVPDKNHHPRAEEAFKILRAAWDIVSNPERRKEYEMKR
                     MAENELSRSVNEFLSKLQDDLKEAMNTMMCSRCCQGKHRRFEMDREPKSARYCAECNRL
                     HPAEEGFWAESSMLGLKITYFALMDGKVYDITEWAGCQRVGISPDTHRVPYHISFGS
                     RVPGTSGRQATPESPPVDLQDFLSRIFQVPPGPMSNGNFFAAPHPGTTSTSRPNS
                     SVPKGEAKPKRKKVRRPFQR"
BASE COUNT      618 a    575 c    698 g    511 t
ORIGIN
```

Query Match   69.2%;   Score 793.6;   DB 10;   Length 2402;
Best Local Similarity 89.2%;   Pred. No. 2.5e-245;
Matches 880; Conservative   0; Mismatches 99; Indels 7; Gaps 2;

```
Qy     1 GCCTTGGGTCAAGCAGAGAATATTAATAGGCAGGGGAATGCACCTGTAGCTAGCTGGCGCTA
         ||| ||||||| |||||||| || ||||  |||| || |||||| |||||||||| ||||
Db  1353 GCCCTGGGTCAAACAGAGGACTAAGAAACAGGGCAATGCACCTGTAGCTAGCTGGCGGCGATA 1412

Qy    61 CTGCCAGCCTGAGAGAGGAAGTGGCTCGACTCTTGACCATGGCTGGGGTTCCTGAGGATGA 120
         || ||||||||||||| |||||| |||||||||||||| ||||||||||||||||||||||
Db  1413 CTGCCAGCCTGAGAGAGGAAGTGACCGGGACTCTTGACCATGGCTGGGGTTCCTGAGGATGA 1472

Qy   121 GCTAAAACCCTTTCCATGGCTACTGGGGGTTGAGGCCACAGCATCAGATGTTGAACTGAAGAA 180
         |||||||||||||||||||| ||||||| ||||||||||||||||||||||||||||||||
Db  1473 ACTAAAACCCTTTTCATGTGCTGCAGTGACTTGAAGCTACAGCATCCGACACACTGAACTAAGAA 1532

Qy   181 GGGCTATAGACAGCTGCAGTGATGGTTCATCCTGACACAAAATCATCATCCCGGGCTGA 240
         ||| ||||| |||||||||||||||| || |||||||||||||| |||||| ||||||||
Db  1533 GGGCCTATAGCCAGCTAGCAGTAATGGTCCATCCTGATAAAAATCACCATCCCCGGGGCTGA 1592

Qy   241 GGAGGGCTTCAAGGTTTTGCAGGCAGCTTGGGACATTGTCAGGAAATGCAAATGCTGAAAAGCGAAA 300
         ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1593 GGAGGGCCTTCAAAAATTTTGCGGGCAGCTTGGGACATTGTCAGCAACTTAACCCAGAGCGAGAGAAGCGAA 1652

Qy   301 GGAGTATGAGATGAAACGAACGGATGGCAGAGAAATGAGCTGAGCCGTCAGTAAATGAGTTTCT 360
         ||||||||||||||||||||| ||||||| |||| |||||||||| |||||||||||||||||
Db  1653 GGAATATGAGATGAAACGGATGAGCAGATGGAGCAGAGAGAAATGAGCTCAGCCGGTCAGTGAATGAGTTTCT 1712

Qy   361 GTCCAAGCTGCGAAGATGACCTCAAGGAGCCAATGAATAACTACTATGATGTGTAGCCGATGCCA 420
         ||| ||| ||||||||||||||||||||||||||||||| ||||||||| ||||||||||||||
Db  1713 GTCGAAACTACACAAGATGAACCTCAAGGAGCCAATGAACAACGACGATGATGTGCAGCAGATGCCA 1772

Qy   421 AGGAAAGCATAGGAGGTTTGAAATGACCTAAGGAGCCCGGAACCTAAGACGTGCTGA 480
         |||||||||||||||||||||||||| ||||||||||||||||||||| ||||||||
Db  1773 AGGAAAGCATAGGAGGTTTGAAATGACCTAAGGAGCCCGGAACCCGGAGCAGAGTACTGTGCTGA 1832

Qy   481 GTGTAATAGGTGCATCCTGTGTGAGAAGGAGACTTTTGGGCAGAGTCAAGCATGTGGG 540
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1833 GTGTAACAGGCTGCATCCTGCTGCTGAGGAAGGAGACTTTTGGGCAGAGTCGAGCATGCTGGG 1892

Qy   541 CCTCAAGATCACCTACTTTGCACTGATGGATGGAAAGTGTATGACATCACAGAGTGGGC 600
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1893 CCTCAAGATCACTTACTTTGCGCTGATGGATGGAAAGTGTATGACATTACAGAGTGGGC 1952

Qy   601 TGGATGCAGCGTGTAGTATCTCCCAGATACCCATCACGAGTCCCTATCACATCTCATT 660
         |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db  1953 TGGATGCAGCGTGTCGTATCTCCCAGATACTCACAGAGTCCTTACCACATCTCATT 2012

Qy   661 TGGTTCTCGGATTCCAGGACCAGAGGGGCGCAGAGAGCCACCCCAGATGCCCTCCTGC 720
         |||||| |||||||||||||||||||||||| || ||||||||||||||||||||||||
Db  2013 TGGTTCTCGGGTACCCGGCACCCGGCCACTCCAGAGTCCCGATGTCCAATGG 2072

Qy   721 TGATCTTCAGGATTCTTGAGTCGGATCTTTCAAGTACCCCCAGGGCAGATGCCCAATGG 780
         ||||||||||| ||||||||||| ||||||||||||||||||| |||||||||||||||
Db  2073 TGACCTGCAGGATTTCTTGAGCCGGATCTTTCAAGTACCTCCGGGGCCGATGTCCAATGG 2132

Qy   781 GAACTTCTTTGCAGCTCCTCAGCTCGCCCGCCCCTGGAGCCGCTGCTTAAGCCCAACAG 840
         |||||||||||||||||||||| ||||||||||| ||||||||||||||||||||||||
Db  2133 GAACTTCTTTGCCGCACCCCACCCTGGCCCTGCCACTCGACCTCCTAGGCCCAACAG 2192

Qy   841 CACAGTACCCAAGGAGAAGCAAACCTAAGCGCGGCGAGAAAGTGAGGAGGCCCTTCCA 900
         |||||||||||||||||||||||| ||||||||||| ||||||||||||||||| ||||
Db  2193 TTCAGTACCCAAGGAGAAGCAAACCTAAACGGCCGGAAGAAAGTGAGGCGGCCGCTTTCA 2252

Qy   901 ACGTTGATGCCCCTTCTCT---TTCCTCAAATCAATGTCAGGGAGTCAAAGGGCTGT-- 955
         ||||||||||||||| ||||   ||||||||||||||||||||||||||||||| ||||
Db  2253 ACGATGACACCCCCTTCTTCTTCCTCTCAAATCAATGTCAGGGAGTCAAAGGGCTGTGT 2312

Qy   956 --AGCACAGGATGGAGTTTGATTTAT 979
           ||||||||||||||||||||||||
Db  2313 ACAGCACCAGGATGGAATTTGATTTAT 2338
```

```
RESULT 12
AF351783
LOCUS       AF351783              2392 bp    mRNA     linear   ROD 24-MAY-2001
DEFINITION  Rattus norvegicus dopamine receptor interacting protein (Drip78)
            mRNA, complete cds.
ACCESSION   AF351783
VERSION     AF351783.1  GI:14194054
KEYWORDS    .
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 2392)
  AUTHORS   Bermak,J.C., Li,M., Bullock,C. and Zhou,Q.Y.
  TITLE     Regulation of transport of the dopamine D1 receptor by a new
            membrane-associated ER protein
  JOURNAL   Nat. Cell Biol. 3 (5), 492-498 (2001)
   MEDLINE   21231375
   PUBMED    11331877
REFERENCE   2  (bases 1 to 2392)
  AUTHORS   Bermak,J.C., Li,M., Bullock,C.M. and Zhou,Q.Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-FEB-2001) Pharmacology, University of California,
            19182 Jamboree Blvd., Irvine, CA 92697, USA
FEATURES             Location/Qualifiers
     source          1..2392
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /note="extrachromosomal"
     gene            1..2392
                     /gene="Drip78"
     CDS             161..2266
```

```
Db   1497 GGAAAGCATAGGAGAGTTTGAAATGACCGGGAACCTAAGAGTGCCAGATACTGTGCTGAG 1556
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    482 TGTAATAGGCTGCATCCTGCTGGAGGAAGGAGACTTTTGGGCAGAGTCAAGCATGTTGGGC 541
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1557 TGTAATAGGCTGCATCCTGCTGGAGGAAGGTGACTTTTGGGCAGAGTCAAGCATGTTGGGC 1616

Qy    542 CTCAAGATCACCTACTTTGCACTGAAGGTATGGATGGAAAGTGTATGACATCACAGAGTGGGCT 601
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1617 CTCAAATCACCTACTTTGCGGCTGATGATGGAAAGTGTATGATATCACAGAGTGGGCT 1676

Qy    602 GGATGCCAGCGTGTAGGTATCTCCCAGATACCCAGAGTCCCTATCACATCTCATT 661
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1677 GGATGCCAGCGTGTGGGAATCTCCCAGATACCCACAGTCCCTTATCACATCTCATT 1736

Qy    662 GGTTCTCGGATTCCAGGCACCACCGAGAGAGCGAGAGAGAGCCACCCAGATGCCCAATCTGCT 721
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1737 GGTTCACGGATGCCAGGCACCACCAGTGGGCGGCGAGAGAGCTACTCCGAGATGCCCCTCCTGCT 1796

Qy    722 GATCTTCAGGATTTCTTGAGTGTCGGATCTTTCAAGTACCCCCAGGGCGCAGATGGCCAATGGG 781
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1797 GACCTTCAGGATTTCTTGAGCCGATCTTTCAAGTACCCCCAGGGCCAGGCCTCCAAGCCAACCAGGG 1856

Qy    782 AACTTCTTTGCAGCTCCTCAGCCTGCCCTGAGCCCGCTGAGCGATGTCTAAGCCACCAGC 841
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1857 AACTTCTTTGCAGCTCCTCAGCCTGCCCTGGGCGCCCCACTGCAGCCTCCAGCCCAACAGC 1916

Qy    842 ACAGTACCCAAGGAGGAAGCCAAACCTAAGCGGCGGAAGAAAGTGAGGAGGCCCTTCCAA 901
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1917 ACAGTACCCAAGGGAGGAAGCCAAACCGAAGCGGCGGAAGAAAGTGAGGAGGCCCTTCCAA 1976

Qy    902 CGTTGATGCCCCTTCCTTCTTTCCTCCAAATCAATGTCAGGGAGTCAAAAGGGCTGT 955
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1977 CGTTGACACCCCTTCTTCTTTCCTCCAATCAATGTCAGGGAGTCAGGGAGTCAAAAGGGCTGT 2030
```

RESULT 11
```
BC011146
LOCUS        BC011146                2402 bp    mRNA    linear    ROD 20-SEP-2002
DEFINITION   Mus musculus, RIKEN cDNA 5730551F12 gene, mRNA, complete cds.
             IMAGE:4016209, mRNA, complete cds.
ACCESSION    BC011146
VERSION      BC011146.1    GI:15029845
KEYWORDS     MGC.
SOURCE       Mus musculus (house mouse)
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1  (bases 1 to 2402)
  AUTHORS    Strausberg,R.
  TITLE      Direct Submission
  JOURNAL    Submitted (25-JUL-2001) National Institutes of Health, Mammalian
             Gene Collection (MGC), Cancer Genomics Office, National Cancer
             Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
             USA
REMARK       NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT      Contact: MGC help desk
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: Gilbert Smith, Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Baylor College of Medicine Human Genome
             Sequencing Center
             Center code: BCM-HGSC
             Web site: http://www.hgsc.bcm.tmc.edu/cdna/
             Contact: amg@bcm.tmc.edu
             Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
             Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
             Richards, S., Gibbs, R.A.
             
             Clone distribution: MGC clone distribution information can be found
             through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
             Series: IRAK Plate: 24 Row: a Column: 19
```

FEATURES   Location/Qualifiers
     source   1..2692
              /organism="Bos taurus"
              /mol_type="mRNA"
              /db_xref="taxon:9913"
     CDS   <1..1982
              /note="bDnaJ1"
              /codon_start=3
              /product="DnaJ1 protein"
              /protein_id="AAG59810.2"
              /db_xref="GI:15843561"
              /translation="SAGSAPNGTRCLTEHSSPXYTQPPNPAHWSDPSHGPPRGPGPPL
              AEEDPDQSEASSESGVDQELSRRENETGYQDDGNSSFLSIPSTCNCQTPGIPEGPYS
              EGRDSSSNFCHHCTSPALGEDKELEGEVDEEEELLSFPVMFHVCPSEKKPAPRQRPR
              VPAKEDTREGGRRDPRSPGRHRLGRKRSQADKRRGLWGAEELCQLGQAGFWWLIEL
              LVLVGENVETGCHLIYACRQLKGSDLDLSRVWVGVWAGRLRGHAQVMFQFLSQGFCYG
              AGLFTRFLRLVGAILLLALALLLGLQLGWRDKATWIFSWLASPT
              WQRCLILLRESRPWQQLVRIVQWGWLELPWVKQRTNRQANAPVAGGRYCQPEEEVARL
              LTMAGVPEDELNPFHVLGVEATASDVELKKAYRQLAVMVHPDKNHHPRAEEAFKVLRA
              AWDIVSNPERRKEYKEVVRAENELSRSVNEFLSKLQEAMNTMMCSRCQGKHRSFEMDR
              EPKSARYCAECNRLHPAEEGDFWAESSMLGLKITYFALMDGKVYDITEWAGCQRVGIS
              PDTHRVPYHISFGSRMPGTSGRQRATPDAPPADLQDFLSRIFQVPPGQMSNGNFFAAP
              QPGPGATAASKPNSTVPVKEAKPXRRKKVRRPFQR"

```
BASE COUNT    670 a    635 c    756 g    631 t
ORIGIN
```

```
Query Match    71.8%;    Score 824;    DB 4;    Length 2692;
Best Local Similarity    92.5%;    Pred. No. 3.5e-255;
Matches 882;    Conservative    0;    Mismatches   60;    Indels   12;    Gaps    1;

Qy      2 CCTTGGGTCAAGCAGAATATTAATAGGCAGGGGAATGCACCTGTAGCTAGTGGGCGCTAC 61
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1089 CCTTGGGTCAAACAGAGAGCCAATAGGCAGGGCGAATGCACCTGTAGCTGGTGTGGTCGTTAC 1148

Qy     62 TGCCAGCCTGAAGAGGAAGTGGCTGACTTGACCTGGCTGGGGTTCCTGAGGATGAG 121
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1149 TGCCAGCCTGAAGAGGAAGTGGCTGACTTCGACTCGGGGTTCCTGAGGATGAG 1208

Qy    122 CTAAACCCTTCCATGTACTGGGGGTTGAGGCCACAGCACAGATGTTGAACTGAAGAAG 181
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1209 CTAAACCCTTTCACGTGTTGGGGGTTGAAGCCACCAGCACATCAGAGTTGAACTGAAGAAG 1268

Qy    182 GCCTATAGACAGCTGGCAGTGATGGTTCATCCTGACAAAAATCATCATCCCCGGCTGAG 241
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1269 GCCTATAGGCAGCTGGCAGTGATGGTTCATCCTGACAAAAATCATCATCCTCGTCGCGGAG 1328

Qy    242 GAGGCCTTCAAGGTTTGCGAGCAGCAGCTTGGACATTGTCAGCAATGCTGAAAGCGAAAG 301
          ||||||||||||||||||||||||||||||||||||||||||||  ||||||||||||||
Db   1329 GAAGCTTTCAAGGTTTGCAAGGTTTTGCGGGCAGCAGCTTGGACATTGTCAGCAACCCTGAAAGACGGAAG 1388

Qy    302 GAGTATGAGATGAGAAACGAATGGCAGAGAATGAGCTGAGCCGGTCAGTAAATGAGTTTCTG 361
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1389 GAATATGAGATGAGAAACGAATGGCAGAGAAAATGAGCTGAGCCGGTCAGTCAATGAGTTTCTG 1448

Qy    362 TCCAAGTTGCAAGATGCCAATGAGCCAATGAAATACTAATGATGTGTAGCCGATGCCAA 421
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1449 TCCAAGTTGCAA------GAAGCAATGAATACGATGATGTGCAGCCGATGCCAG 1496

Qy    422 GGAAAGCATAGGAGAGTTTGAAATGACCGGGAACCTAAGAGTGCCAGATACTGTGCTGAG 481
```

```
Db   1574  GCCTATAGGCAGCTGGCAGTGATGGTTCATCCTGACAAAATCATCATCCTCGTGCTGAG  1633
                  ||||||||||||||||||||||||||||||| ||||||||  |||||| |||||||||||
Qy    242  GAGGCCCTCAAGTTTTGCGAGCAGCTGGGCACATTGTCAGCAATGCTGAAAAGCGAAAG   301

Db   1634  GAAGCCCTCAAGTTTTGCGGGCAGCTGGGACATTGTCAGCAACCCTGAAAGACGGAAG   1693
                  ||||||||||||||||||| |||||||| ||||||||||||||||  |||||  |||||
Qy    302  GAGTATGAGATGAAACGAACGAATGGCAGAGAGGCGAGATGAGCTAGTAAATGAGTTCTG   361

Db   1694  GAATATGAGATGAAACGAATGGCAGAGAGAAAATGAGCTAGTCAGTGAATGAGTTCTG   1753
                  ||||||||||||||||||| ||||||||  ||||||||||| |||||||||||||||||
Qy    362  TCCAAGTGCAAGATGACCTCAAGGAGGAGCAATGAATACTATGATGTGTAGCCGATGCCAA   421

Db   1754  TCCAAGCTGCAA--------GAAGCAATGAATACGATGATGTGCAGCCGATGCCAG   1801
                  ||||| |||||          |||| |||||||| ||||||||  |||||||||||||
Qy    422  GGAAAGCATAGGAGGTTTGAAATGACCGGGAACCTAAGAGTGCCAGATACTGTGCTGAG   481

Db   1802  GGAAAGCATAGGAGGTTTGAAATGACCGGGAACCTAAGAGTGCCAGATACTGTGCTGAG   1861
                  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    482  TGTAATAGGCTGCATCCTGCTGAGGAAGGAGACTTTGGGCAGAGTGACTTTGGGCAGCAT   541

Db   1862  TGTAATAGGCTGCATCCTGCTGAGGAAGGTGACTTTGGGCAGAGTGACTTTGGGC   1921
                  ||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Qy    542  CTCAAGATCACCTACTTTGCACTGATGATGGAAAGGTGTATGACATCACAGAGTGGGCT   601

Db   1922  CTCAAATCACCTACTTTGCGGTGATGATGGAAAGGTGTATGATATATCCACAGAGTGGGCT   1981
                  ||||| ||||||||||||| |||||||||||||| |||||||| ||| ||||||||||||
Qy    602  GGATGCCAGCGTGTAGGTATCTCCCAGATACCCACAGAGTCCCTATCACATTCCATTT   661

Db   1982  GGATGCCAGCGTGTGGGAATCTCCCAGATACCCACAGAGTCCCTATCACATTCCATTT   2041
                  ||||||||||||||||| | ||||||||||||||||||||||||||||||||||||
Qy    662  GGTTCGGATTCCAGGCACCAGGCGGCGCACCCGAGATGCCCCTCCTGCT   721

Db   2042  GGTTCACGGATGCCAGGCACCAGTGGGCGCGGAGAGCTACTCCAGATGCCCCTCCTGCT   2101
                  ||||| ||||| |||||||||| ||| ||||| |||||||| ||||||||||||||||
Qy    722  GATCTTCAGGATTTCTTGAGTCGGATCTTCAAGTACACCCCAGGGCAGCCCAATGGG   781

Db   2102  GACCTTCAGGATTTCTTGAGCCGGATCTTCAAGTACACCCCAGGCCAGCCCAACGGG   2161
                  || |||||||||||||||| |||||||||||||||||||||||| || ||||| ||||
Qy    782  AACTTCTTGCAGCTCCTCAGCCTGCCCCTGCAGCTCCTCAAGCCCAACAGC   841

Db   2162  AACTTCTTGCAGCTCCTCAGCCTGCCCCTGGGCCCACTGCAGCTCCTCCAAGCCCAACAGC   2221
                  |||||||||||||||||||||||||||||  |||| ||||||||||| |||||||||||
Qy    842  ACAGTACCCAAGGGAGAAGCCAAACCTAAGCCGCGGCAGAAGAAAGTGAGGAGCGCCCTTCAA   901

Db   2222  ACAGTACCCAAGGGAGAAGCCAAACCGCGGCAGAAGAAAGTGAGGAGCGCCCTTCCAA   2281
                  ||||||||||||||||||||||||||  |||||||||||||||||||||||||||||  ||
Qy    902  CGTTGATGCCCCTTCCTCTTCCCTCAAATCAATGTCAGGGAGTCAAAAGGGCTGT----AG   957

Db   2282  CGTTGACACCCCCTTCTCTTCCTCCAATCAATGTCAGGGAGTCAAAAGGGCTGTGTACAG   2341
                  |||||| | ||||  |||||||| ||||||||||||||||||||||||||||||||
Qy    958  CACAGGATGGAGTTGATTAT   979

Db   2342  CACCAGAGTGGAGTTTGATTTGT   2363
```

RESULT 10
AF308815
LOCUS   AF308815   2692 bp   mRNA   linear   MAM 18-JAN-2002
DEFINITION Bos taurus DnaJ1 protein mRNA, partial cds.
ACCESSION   AF308815
VERSION   AF308815 2 GI:15843560
KEYWORDS   .
SOURCE   Bos taurus (cow)
  ORGANISM Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1 (bases 1 to 2692)
  AUTHORS   Neill,J.D. and Ridpath,J.F.
  TITLE   Recombination with a cellular mRNA encoding a novel DnaJ protein
            results in biotype conversion in genotype 2 bovine viral diarrhea
            viruses
  JOURNAL   Virus Res. 79 (1-2), 59-69 (2001)

ACCESSION   AY027881
VERSION   AY027881.1 GI:15777192
KEYWORDS   .
SOURCE   Bos taurus (cow)
  ORGANISM Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1 (bases 1 to 3159)
  AUTHORS   Rinck,G., Birghan,C., Harada,T., Meyers,G., Thiel,H.J. and Tautz,N.
  TITLE   A cellular J-domain protein modulates polyprotein processing and
            cytopathogenicity of a pestivirus
  JOURNAL   J. Virol. 75 (19), 9470-9482 (2001)
  MEDLINE   21424530
  PUBMED   11533209
REFERENCE   2 (bases 1 to 3159)
  AUTHORS   Rinck,G., Tautz,N. and Meyers,G.
  TITLE   Direct Submission
  JOURNAL   Submitted (20-FEB-2001) Institut fuer Virologie (FB10),
            Justus-Liebig-Universitaet Giessen, Frankfurterstrasse 107, Giessen
            35392, Germany
FEATURES   Location/Qualifiers
     source   1..3159
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
     5'UTR   1..187
     CDS   188..2287
                     /function="interacts with nonstructural protein 2 of
                     bovine viral diarrhea virus"
                     /note="contains putative zinc finger domain"
                     /codon_start=1
                     /product="J-domain protein Jiv"
                     /protein_id="AAK28650.1"
                     /db_xref="GI:15777193"
                     /translation="MAQKHPGEGGLCGAHHSGGASLRTLGPSVDPDILSFSGLRDSAG
                     SAPNGTRCLTEHSSPKYTQPPNPAHWSDSHGPPRGPGGPPLAEEDPDQSEASSEESGV
                     DQELSRENETGYQDDGNSSFLSIPSTCNCQGTPGIPEGPYSEGRDSSSSNFCHHCTSP
                     ALGEDEELEGEYDEEEPLKFPSDVSRVPSEKKPAPRRQRHRVPAKEDTREGGRRDPRS
                     PGRHRLGRKRSQADKRRGLGLWGAEELCQLGQAGFWLIELLVLVGEVVETCGHLIYA
                     CRQLKGSDLDLLRVWVGVWAGRLRGWAQVMFQFLSQGFCYGAGLFTRFLRIVGALLLL
                     ALALLLGCLQLGWRFLVGLSDRLGWRDKATWIFSWLASPTWQRCLILRESRPWQQLV
                     RIVQWGWLELPWVXQRTNRQANAPVAGGRYCQPEEEVARLLTMAGVPEDELNPPHVLG
                     VEATASDVELKKAYRQLAVMVHPDKNHHPRAEEAFKVLRAAWDIVSNPERRKEYEMKR
                     MAENELSRSVNEFLSKLQEAMNTMMCSRCQGKHRRFEMDREPKSARYCAECNRLHPAE
                     EGDFWAESSMLGLKITYFALMDGKVYDITEWAGCQRVGISPDTHRVPVHISFGSRMPG
                     TSGRQRATPDAPPADLQDPLSRIFQVPPQOMSNGNFFRAPQPGPGATAASKPNSTVPK
                     GEAKPKRKKVRRPFQR"
     3'UTR   2288..3159
                     /note="use of alternative polyadenylation site results in
                     3'UTR longer than in the sequence presented in GenBank
                     Accession Number AY027882"
BASE COUNT   755 a   759 c   895 g   750 t
ORIGIN

Query Match   72.7%; Score 834; DB 4; Length 3159;
Best Local Similarity   92.3%; Pred. No. 2e-258;
Matches 906; Conservative   0; Mismatches 60; Indels 16; Gaps   2;

Qy   2 CCTTGGTCAAGCAGCAGAATATTAATAGGCAGGGAATGCACCTGTAGCTAGTGGGCGCTAC   61
           ||||||||||||||||||||||| ||||||||| |||||||||||||||||||| |||
Db   1394 CCTTGGTCAAACAGGAGACCAATAGGCAGGCAGCGAATGCACCTGTAGCTAGTGGTCGTTAC 1453

Qy   62 TGCCAGCGCTGAAGAGGAAGTGGCTGGCTGACTCTTGACCATGGCTGGGGTTCCTGAGGATGAG 121
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1454 TGCCAGCGCTGAAGAGGAAGTGGCTGGCTGACTCTTGACCATGGCTGGGGTTCCTGAGGATGAG 1513

Qy   122 CTAAACCCTTCCATGTACTGGGGTTGAGGCCACAGCATCAGATGTTGAACTGAAGAAG 181
           |||||||| ||||||||||||||||||| || ||||||||||||||||||||||||||||
Db   1514 CTAAACCCTTTCACGTGTTGGGGGGTTGAAGCCACAGCCATCAGATGTTGAACTGAAGAAG 1573

Qy   182 GCCTATAGACAGCTGGCAGTGATGGTTCATCCTGACAAAATCATCATCCTCGTGCTGAG 241
           |||||| || |||||||||||||||||||||||||||||||||||||| |||||||
```

Qy      2  CCTTGGGTCAAGCAGAGAATATTAATAGGCAGGGAATGCACCTGTAGTGGGCGCTAC    61
           |||||||||||||||| |||||||||| |||||||||||||||||||||||||||||
Db   1394  CCTTGGGTCAAACAGAGAGACCAATAGGCAGGGAATGCACCTGTAGTGGTGTCGTTAC  1453

Qy     62  TGCCAGCTGAAGAGGAAGTGGCTGACTCTTGACCATGTGACCTGGGGTTCCTGAGGATGAG   121
           |||||||||| |||||  |||||||||||||||||||||| |||||||||||||||||||||
Db   1454  TGCCAGCCTGAACAGCTGGCAGTGGTCGCACTCGTGACCTCTGACCATGTCCTGAGGATGAG  1513

Qy    122  CTAAACCTTCCATGTACTGGGGTTGAGGCCACAGCATCAGATGTTGAACTGAAGAAG   181
           ||| |||||| |||||||||||||||||| ||||||||||| ||||||||||||||||
Db   1514  CTAAACCCTTTCACGTTGTTGGGGGTTGAAGCCACAGCATCAGATGTTGAACTGAAGAAG  1573

Qy    182  GCCTATAGACAGTGGCAGTGATGGTTCATCCTGACAAAAATCATCCCGGCTGAG   241
           ||||||||||||||||||| ||||||||||||||||  |||||||||||||||||||
Db   1574  GCCTATAGGCAGTGGCAGTGGTCATGGTTCATCCTGACAAAAATCATCCCTGTGCTGAG  1633

Qy    242  GAGGCCTTCAAGGTTTGCGAGCAGTTGGGACAGTTGAGGCCACAGCATTGTCAGCAATGTGAAAAGCGAAAG   301
           ||||||||||||||||||||||||| |||||||| ||||| ||||||||||||
Db   1634  GAAGCCTTCAAGGTTTGCGAGCAGTTGGGGCAGTTGCAGCAGCACCCTGCTGCAGCGGAAG  1693

Qy    302  GAGTATGAGATGAAACGAATGGCGAGAGAATGACTGAGCGCCGGTCAGTAAATAGTTTCTG   361
           ||| ||||| || |||||||||||| ||||||||||||| |||||||||||||||||||
Db   1694  GAATATGAGATGAAACGAATGGCAGAAATGACTGAGCGCCGGTCAGTGAATGAGTTTCTG  1753

Qy    362  TCCAAGCTGCAAGATGACCTCAAGGAGCAATGAATACTATGATGTGTAGCCGATGCCAA   421
           |||||||||||            |||||| || |||||||||||||||||||||
Db   1754  -------TCCAAGCTGCAA---------GAAGCAATGAATACGATGATGTGCAGCGCCAGATGCCAG  1801

Qy    422  GGAAAGCATAGGAGGTTTGAAATGGACCGGGAACCTAAGAGTGCCAGATACTGTGCTGAG   481
           ||||||||||||||||||||||| |||| ||||||||||||||||||| |||||||||||||
Db   1802  GGAAAGCATAGGAGGTTTGAAATGGACCGGGAACCTAAGAGTGCCAGATACTGTGCTGAG  1861

Qy    482  TGTAATAGGCTGCATCCTGCTGAGGAAGGAGGTGACTTTTGGGCAGAGTCAAGCATGTTGGGC   541
           ||||||||||||||||||||||||||||||||| ||||||| ||||||||||||||||||||||
Db   1862  TGTAATAGGCTGCATCCTGCTGAGGAAGGTGACTTTTGGGCAGAGTCAAGCATGTTGGGC  1921

Qy    542  CTCAAGATCACCTACTTGCACTGATGGATGAAAAGTGTATGACATCACAGAGTGGGCT   601
           ||| |||| |||||||||||||||||| |||||||||||| ||||||| ||||||||||||||
Db   1922  CTCAAAATCACCTACTTTGCCTGATGGATGAAAAGTGTATGATATCACAGAGTGGGCT  1981

Qy    602  GGATGCCAGCGTGTAGTATCCCCAGATACCCACAGAGTCCCTATCACATCTCATTT   661
           ||||||||||| |||||||||||||||||||||| ||||||||||| ||||||||||||
Db   1982  GGATGCCAGCGTGTGGGAATCTCCCCAGATACCCACAGAGTCCCTTATCACATCTCATTT  2041

Qy    662  GGTTCTCGGATTCCAGGCACCAGCACACCGTGGCGCCGAGAGCCACCCCAGTGCCCTCCTGCT   721
           ||||| ||||||||||||| ||||||||| ||||||||||| |||||| |||||||||||||||
Db   2042  GGTTCACGGATGCCAGGCACCAGCACCCGTGGGCGCCAGAGAGCACTCCAGTGCCCTCCTGCT  2101

Qy    722  GATCTTCAGGATTCTTGAGTCGGATCTTTCAAGTACCCCCAGGGCAGATGCCAATGGG   781
           |||||||||||||||||||||| |||| ||||||||||||||| ||||||||||||||||||
Db   2102  GACCTTCAGGATTCTTGAGCCGGATCTTTCAAGTACCCCCAGGCCGCCAGATGTCCAACGGG  2161

Qy    782  AACTTCTTGCAGCTCTCAGCCTGCCCCTGAGCCTCTAAGCCCAACAGC   841
           ||||||||||||||||||| |||||||||||||||| |||||||||||||||
Db   2162  AACTTCTTGCAGCTCTCAGCCTGGGCCCTGGGCACTGCGAGCCTCCAAGCCCAACAGC  2221

Qy    842  ACAGTACCCAAGGGAAGCCAAACTTGAGGAAGAAGTGACGGAGGCCCTTCCAA   901
           |||||||||||||||||||||||||||| ||||||||||||| ||||||||||||||||||
Db   2222  ACAGTACCCAAGGGAAGCCAAACCCAAGCGGCGGAAGAAAGTGAGGAGGAGGCCCTTCCAA  2281

Qy    902  CGTTGATGCCCCTTCTTCTTTTTTGATTTAT----AG   957
           |||| |||||||||||||||||||||||||||   ||
Db   2282  CGTTGACACCCCTTCTTCTTTTCTCCAATCAATGTCAGGGAGTCAAAAGGGCTGT----AG  2341

Qy    958  CACAGGATGGAGTTTGATTTAT   979
           ||||||||||||||||||||||
Db   2342  CACCAGGATGGAGTTTGATTTGT  2363

Qy   1042  GCTCGTTGCCCAGCTGGAGTGGCAGTGGTGTGATCTCAGCTTACTGCAACCTCTGTCTCC  1101
           |||||||||||||||||||||||||||||  |||||| |||||||||||||||||||||||||
Db    781  GCTCGTTGCACAGCTGGAGTGGCAGTGGTGTGATCTCAGCTTACTGCAACCTCTGTCTCC   840

Qy   1102  CGGGTTCAAGCAATTCCCATCTCAGCTCCTGAGTAGCTGGGAT  1147
           ||||||||||||||||||||||||||||||||||||||||||||
Db    841  CGGGTTCAAGCAATTCCCCATCTCCAGCTCCTGAGTAGCTCGGGAT   886

RESULT 8
AY027882
LOCUS       AY027882         2402 bp   mRNA   linear   MAM 25-SEP-2001
DEFINITION  Bos taurus J-domain protein Jiv mRNA, complete cds.
ACCESSION   AY027882
VERSION     AY027882.1   GI:15777194
KEYWORDS    .
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 2402)
  AUTHORS   Rinck,G., Birghan,C., Harada,T., Meyers,G., Thiel,H.J. and Tautz,N.
  TITLE     A cellular J-domain protein modulates polyprotein processing and
            cytopathogenicity of a pestivirus
  JOURNAL   J. Virol. 75 (19), 9470-9482 (2001)
  MEDLINE   21424530
   PUBMED   11533209
REFERENCE   2  (bases 1 to 2402)
  AUTHORS   Rinck,G. and Tautz,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-FEB-2001) Institut fuer Virologie (FB10),
            Justus-Liebig-Universitaet Giessen, Frankfurterstrasse 107, Giessen
            35392, Germany
FEATURES             Location/Qualifiers
     source          1..2402
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /cell_line="MDBK"
     5'UTR           1..187
     CDS             188..2287
                     /function="interacts with nonstructural protein 2 of
                     bovine viral diarrhea virus"
                     /note="contains putative zinc finger domain"
                     /codon_start=1
                     /product="J-domain protein Jiv"
                     /protein_id="AAK28651.1"
                     /db_xref="GI:15777195"
                     /translation="MAQXHPGEGGLCGAHHSGGASLRTLGPSVDPDILSFSGLRDSAG
                     SAPNGTRCLTEHSSPKYTQPPNPAHWSDPSHGPPRGPGPPLAEEDPDQSEASSEESGV
                     DQELSRENTGYQDDGNSSFLSIPSTCNCQGTPGIPEGPYSEGRDSSSNFCHHCTSP
                     AIGEDEELEGEYDEEEPLKFPSDVSRVPSEKKPAPRRQRHRVPAKEDTREGGRRDPRS
                     PGRHRLGRKRSQADKRRGLGLWGAEELCQLGQAGFWWLIELLVLVGEYVETCGHLIYA
                     CRCLKGSDLDLLRVWVGVWAGRLRGWAQVMFQFLSQGFCYGAGLFTRFLRLVGALLL
                     AIALLLGCLQLGWRFLVGLSDRLGWRDKATWIFSWLASPTWQRCLILLRESRPWQQLV
                     RIVQWGWDELPWVKQRTNRQANAPVAGGRYCQPEEEVARLLTMAGVPEDELNPFHVLG
                     VEATASDVELKKAYRQLAVMVHPDKNHHPRAEEAFKVLRAAWDIVSNPERRKEYEMKR
                     MAENELSRSVNEFLSKLQEAMNTMMCSRCQGKHRRFEMDREPKSARYCAECNRLHPAE
                     EGDFWAEESSMLGLKITTYFALMDGKVYDITEWAGCQRVGISPDTHRVPYHISFGSRMPG
                     TSGRQRATPDAPPADLQDFLSRIFQVPFGQMSNGNFFAAPQPGPGATAASKPNSTVPK
                     GEAKPKRKKVRRPFQR"
     3'UTR           2288..2402
                     /note="use of alternative polyadenylation site results in
                     3'UTR shorter than in the sequence presented in GenBank
                     Accession Number AY027881"
BASE COUNT      559 a    582 c    734 g    527 t
ORIGIN

Query Match     72.7%;  Score 834;  DB 4;  Length 2402;
Best Local Similarity   92.3%;
Matches  906;  Conservative   0;  Mismatches  60;  Indels  16;  Gaps   2;

RESULT 9
AY027881
LOCUS       AY027881         3159 bp   mRNA   linear   MAM 25-SEP-2001
DEFINITION  Bos taurus J-domain protein Jiv mRNA, complete cds.

```
                                    /organism="Homo sapiens"
                                    /mol_type="mRNA"
                                    /db_xref="taxon:9606"
                                    /chromosome="12"
CDS              <1..644
                                    /note="DRIP78; ER-membrane protein; zinc finger motif"
                                    /codon_start=3
                                    /product="dopamine receptor interacting protein"
                                    /protein_id="AAK56241.1"
                                    /db_xref="GI:14194057"
                                    /translation="AWDIVSNAEKRKEYEMKRMAENELSRSVNEFLSKLQDDLKEAMN
                                    TMMCSRCQGHRREFMDREPKSARYCAECNRLHPAEEGDFWAESSMGLKITYFALMD
                                    GKVVDITEWAGCQRVGISPDTHRVYHISFGSRIPGTRGRQRATPDAPPADLQDFLSR
                                    IFQVPPGQMPNGNFFAAPQPAPGAAAASKPNSTVPKGEAKPKRKKVRRPFQR"

BASE COUNT       240 a     236 c     246 g     231 t
ORIGIN

Query Match          75.6%;   Score 867.2;   DB 9;   Length 953;
Best Local Similarity 99.4%;   Pred. No. 3.1e-269;
Matches    881;   Conservative    0;   Mismatches    0;   Indels    2;   Gaps    1;

Qy  264  CAGCTTGGGACATTGTCAGCAATGCTGAAAAGCGAAAGGAGTATGAGATGAAACGAATGG  323
         |||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db    1  CAGCTTGGGACATTGTCAGCAATGCTGAAAAGCGAAAGGAGTATGAGATGAAACGAATGG   60

Qy  324  CAGAGAATGAGCTGAGCCGGTCAGTAAATGAGTTTCTGTCCAAGCTGCAAGATGACCTCA  383
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  CAGAGAATGAGCTGAGCCGGTCAGTAAATGAGTTTCTGTCCAAGCTGCAAGATGACCTCA  120

Qy  384  AGGAGGCAATGAATACTATGATGTGTAGCCGATGCCAAGGAAAGCATAGGAGGTTTGAAA  443
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  AGGAGGCAATGAATACTATGATGTGTAGCCGATGCCAAGGAAAGCATAGGAGGTTTGAAA  180

Qy  444  TGGACCGGGAACCTAAGAGTGCCAGATACTGTGCTGAGTGTAATAGGCTGCATCCTGCTG  503
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  TGGACCGGGAACCTAAGAGTGCCAGATACTGTGCTGAGTGTAATAGGCTGCATCCTGCTG  240

Qy  504  AGGAAGGAGACTTTGGGCAGAGTCAAGCATGTTGGGCCTCAAGATCACCTACTTTGCAC   563
         |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db  241  AGGAAGGAGACTTTGGGCAGAGTCAAGCATGTGGGGCCTCAAGATCACCTACTTTGCAC  300

Qy  564  TGATGGATGGAAAGGTGTATGACATCACAGAGTGGGCTGAGTGCCAGCGTAGGTATCT   623
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301  TGATGGATGGAAAGGTGTATGACATCACAGAGTGGGCTGAGTGCCAGCGTAGGTATCT  360

Qy  624  CCCCAGATACCCCACAGAGTCCCCTATCACATCCATTGGTTCTCGGATTCCAGGCACCA  683
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361  CCCCAGATACCCCACAGAGTCCCCTATCACATCCATTGGTTCTCGGATTCCAGGCACCA  420

Qy  684  GAGGGGCGGCAGAGAGCCACCCCAGATGCCCCTCCTGCTCAGCGTTCAGGATTTCTTGAGTC  743
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421  GAGGGGCGGCAGAGAGCCACCCCAGATGCCCCTCCTGCTCAGCGTTCAGGATTTCTTGAGTC  480

Qy  744  GGATCTTTCAAGTACCCCCAGGGCAGATGGGAACTTCTTTGCAGCTCCTCAGC  803
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481  GGATCTTTCAAGTACCCCCAGGGCAGATGGGAACTTCTTTGCAGCTCCTCAGC  540

Qy  804  CTGCCCTGGAGCCGCTGCAGCCTCTTAAGCCCAACAGCACCAGTACCCAAGGGAGAAGCCA  863
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541  CTGCCCTGGAGCCGCTGCAGCCTCTTAAGCCCAACAGCACCAGTACCCAAGGGAGAAGCCA  600

Qy  864  AACCTAAGCGGCGGAAGAAAGTGAGGAGGCCCTTCCAACGTTGATGCCCCTTCTCTTTCC  923
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601  AACCTAAGCGGCGGAAGAAAGTGAGGAGGCCCTTCCAACGTTGATGCCCCTTCTCTTTCC  660

Qy  924  TCAAATCAATGTCAGGGAGTCAAAGGGCTGTAGCACCAGGATGCAAAGGGCTGAGTTGATTTATCCCT  983
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  661  TCAAATCAATGTCAGGGAGTCAAAGGGCTGTAGCACCAGGATGCAAAGGGCTGAGTTGATTTATCCCT  720

Qy  984  CCTCCCCAACACCTAGGAACTGAATCTTTT--CTTTTATTTTTTGAGATGGAGTCTT  1041
         |||||||||||||||||||||||||||||  |||||| |||||||||||||||||||
Db  721  CCTCCCCAACACCTAGGAACTGAATCTTTTTCTTTTTATTTTTTTAATTTTTGAGATGGAGTCTT  780
```

```
Db  271  GGCCTATAGACAGCTGGCAGTGATGTTCATCCTGACAAAATCATCCCGGGCTGA  330

Qy  241  GGAGGCCTTCAAGGTTTGCGAGCAGCTGGGACATTGTCAGCAATGCTGAAAAGCGAAA  300
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  331  GGAGGCCTTCAAGGTTTGCGAGCAGCTGGGACATTGTCAGCAATGCTGAAAAGCGAAA  390

Qy  301  GGAGTATGAGATGAAACGAATGGCAGAGAATGAGCTGAGCCGGTCAGTAAATGAGTTTCT  360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  391  GGAGTATGAGATGAAACGAATGGCAGAGAATGAGCTGAGCCGGTCAGTAAATGAGTTTCT  450

Qy  361  GTCCAAGCTGCAAGATGACCTCAAGGAGGCAATGAATACTATGATGTGTAGCCGATGCCA  420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  451  GTCCAAGCTGCAAGATGACCTCAAGGAGGCAATGAATACTATGATGTGTAGCCGATGCCA  510

Qy  421  AGGAAAGCATAGGAGGTTTGAAATGGACCGGGAACCTAAGAGTGCCAGATACTGTGCTGA  480
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  511  AGGAAAGCATAGGAGGTTTGAAATGGACCGGGAACCTAAGAGTGCCAGATACTGTGCTGA  570

Qy  481  GTGTAATAGGCTGCATCCTGCTGCTGAGGAAGGAGACTTTGGGCAGAGTCAAGCATGTGGG  540
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  571  GTGTAATAGGCTGCATCCTGCTGCTGAGGAAGGAGACTTTGGGCAGAGTCAAGCATGTGGG  630

Qy  541  CCTCAAGATCACCTACTTTGCACTGATGATGAAAGGTGTATGACATCACAGAGTGGGC  600
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  631  CCTCAAGATCACCTACTTTGCACTGATGATGAAAGGTGTATGACATCACAGAGTGGGC  690

Qy  601  TGGATGCCACGTGTAGGTATCTCCCCAGATACCCCACAGAGTCCCCTATCACATCCATT  660
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  691  TGGATGCCACGTGTAGGTATCTCCCCAGATACCCCACAGAGTCCCCTATCACATCCATT  750

Qy  661  TGGTTCTCGGATTCCAGGCACCACCAGAGGGCGGCACCAGAGAGCCACCCCAGATGCCCCTGC  720
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  751  TGGTTCTCGGATTCCAGGCACCACCAGAGGGCGGCACCAGAGAGCCACCCCAGATGCCCCTGC  810

Qy  721  TGATCTTTCAGGATTTCTTGAGTCGGATCTTTCAGGATTTCTTCAAGTACCCCAATGG  780
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  811  TGATCTTTCAGGATTTCTTGAGTCGGATCTTTCAGGATTTCTTCAAGTACCCCAATGG  870

Qy  781  GAACTTCTTTGCAGCTCCTCAGCCTGCCCTGGAGCCGCTGCAGCCTCTTAAGCCCAACAG  840
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  871  GAACTTCTTTGCAGCTCCTCAGCCTGCCCTGGAGCCGCTGCAGCCTCTTAAGCCCAACAG  930

Qy  841  CACCAGTACCCAAGGGAGAAGCCAAACCTAAGCGGCGGCGGAAGAAAGT  885
         ||||||||||||||||||||||||||||||||||||||||||||||| 
Db  931  CACCAGTACCCAAGGGAGAAGCCAAACCTAAGCGGCGGCGGAAGAAACT  975
```

RESULT 7
AF351784

LOCUS   AF351784   953 bp   mRNA   linear   PRI 24-MAY-2001
DEFINITION Homo sapiens dopamine receptor interacting protein mRNA, partial
            cds.
ACCESSION   AF351784
VERSION   AF351784.1   GI:14194056
KEYWORDS   .
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 953)
  AUTHORS   Bermak,J.C., Li,M., Bullock,C. and Zhou,Q.Y.
  TITLE   Regulation of transport of the dopamine D1 receptor by a new
            membrane-associated ER protein
  JOURNAL   Nat. Cell Biol. 3 (5), 492-498 (2001)
  MEDLINE   21231375
   PUBMED   11331877
REFERENCE   2 (bases 1 to 953)
  AUTHORS   Bermak,J.C., Li,M., Bullock,C.M. and Zhou,Q.-Y.
  TITLE   Direct Submission
  JOURNAL   Submitted (21-FEB-2001) Pharmacology, University of California,
            19182 Jamboree Blvd. Irvine, CA 92697, USA
FEATURES   Location/Qualifiers
     source   1..953

```
SOURCE        Homo sapiens (human)
ORGANISM      Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1 (bases 1 to 1818)
AUTHORS       Strausberg,R.
TITLE         Direct Submission
JOURNAL       Submitted (05-NOV-2001) National Institutes of Health, Mammalian
              Gene Collection (MGC), Cancer Genomics Office, National Cancer
              Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
              USA
REMARK        NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT       Contact: MGC help desk
              Email: cgapbs-r@mail.nih.gov
              Tissue Procurement: ATCC
              cDNA Library Preparation: Life Technologies, Inc.
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Sequencing Group at the Stanford Human Genome
              Center, Stanford University School of Medicine, Stanford, CA 94305
              Web site:         http://www-shgc.stanford.edu
              Contact:     (Dickson, Mark) mcd@paxil.stanford.edu
              Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
              R. M.

              Clone distribution: MGC clone distribution information can be found
              through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
              Series: IRAK Plate: 20 Row: a Column: 13
              This clone was selected for full length sequencing because it
              passed the following selection criteria: Similarity but not
              identity to protein.

FEATURES      Location/Qualifiers
     source   1..1818
              /organism="Homo sapiens"
              /mol_type="mRNA"
              /db_xref="taxon:9606"
              /clone="MGC:21452 IMAGE:3448846"
              /tissue_type="Placenta, choriocarcinoma"
              /clone_lib="NIH_MGC_10"
              /lab_host="DH10B"
              /note="Vector: pCMV-SPORT6"
     CDS      401..1570
              /codon_start=1
              /product="Unknown (protein for MGC:21452)"
              /protein_id="AAH16941.1"
              /db_xref="GI:16877384"
              /translation="MKRMAENELSRSVNEFLSKLQDDLKEAMNTMMCSRCQGKHRRFE
              MDREPKSARYCAECNRLHPAEEGDFWAESMLGLKITYFALMDGKVYDITEWAGCQRV
              GISPDTHRVPYHISFGSRIPGTRGRQRATPDAPPADLQDFLSRIFQVPPGQMPNGNFF
              AAPQPAPGAAAASKPNSTVPKGEAKPKRKKLAELKQECLARGLETKGIKQDLIHRLQ
              AYLEHAEEAANEDVLGDETEEEETKPIELPVKEEPPEKTVDVAAEKKVVKITSEI
              PQTERMQKRAERFNVPVSLESKKVARAARFGISSVPTKGLSSDNKPMVNLDKLKERAQ
              RFGLNVSSISRKSEDDEKLKKRKERFGIVTSSAGTGTEDTEAKKRKRAERFGIA"
BASE COUNT     532 a    373 c    494 g    419 t
ORIGIN

Query Match             77.0%;  Score 883.4;  DB 9;  Length 1818;
Best Local Similarity   99.9%;  Pred. No. 1.9e-274;
Matches  884;  Conservative   0;  Mismatches   1;  Indels   0;  Gaps   0;

Qy         1 GCCTTGGGTCAAGCAGAGAATATTAATAGGCAGGGGAATGCACCTGTAGCTAGTGGGCGCTA 60
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        91 GCCTTGGGTCAAGCAGAGAATATTAATAGGCAGGGGAATGCACCTGTAGCTAGTGGGCGCTA 150

Qy        61 CTGCCAGCCTGAAGAGGAAGAGGTGGCTCGACTCTTGACCATGGCTGGGGTTCCTGAGGATGA 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       151 CTGCCAGCCTGAAGAGGAAGAGGTGGCTCGACTCTTGACCATGGCTGGGGTTCCTGAGGATGA 210

Qy       121 GCTAAACCCTTCCATGTACCTGGGGTTGAGGGCCACCAGCATCAGATGTTGAACTGAAGAA 180
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       211 GCTAAACCCTTCCATGTACCTGGGGTTGAGGGCCACCAGCATCAGATGTTGAACTGAAGAA 270

Qy       181 GGCCTATAGACAGCTGGCAGTGATGTTCATCCTGACAAAAATCATCATCCCGGCCTGA 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Qy       241 GGAGGCCTTCAAGGTTTTGCGAGCAGCATTGTCAGCAATGCTGAAAAGCGAAA 300
             ||||||||||||||||||||||                    |||||||||||
Db    117982 GGAGGCCTTCAAGGTTTTGCGCGCAGTTTGGGACAGTTTGCAGCAATGCTGAAAAGCAGAA 118041

Qy       301 GGAGTATGAGATGAAACGAATGCAGAGAATGAGCTGAGCGGTCAGTAAATGAGTTTCT 360
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    118042 GGAATATGAGATGAAACGAATGCAGAGAATCAGCTGAGCCGGTCAGTAAATGAGTTTCT 118101

Qy       361 GTCCAAGCTGCAAGATGACCTTCAAGGACGCAATGAATACTATGATGTGTAGCCGATGCCA 420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    118102 GTCCAAGCTGCAAGATGACCTTCAAGGACGCAGTGAATACTCTGATGTGTGGCCGGATGCCA 118161

Qy       421 AGGAAAGCATAGGAGGTTTGAAATGGACCGGGAACCTAAGAGTGCCAGATACTGTGCTGA 480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    118162 AGGAAAGCATAGGAGGTTTGAGGTGACCGGGAACCTAAGAGTGCCAGATACTATGCTGC 118221

Qy       481 GTGTATAGGCTGCATCCTGCTGAGGAAGGAGACTTTGGGCAGCAGTTTGGGCAAGCATGTGGG 540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    118222 GTGTGATAGGCTGCGTGTCCCGTGAGGAAGGAGACTTTTGGGCAGAGTCAAGCTTGTTGGG 118281

Qy       541 CCTCAAGATCACCTACTTTGCACTGATGGATGGAAAAGTGTATGACATCACAGAGTGGGC 600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    118282 CCTCAAGATCACCTACTTTGCACTGATGGATGGAAAGGTGTATGACATCACACAGTGGGC 118341

Qy       601 TGGATGCCAGCGTGTAGGTAGTATCTCCCCAGATACCCACGAGTCCCTATCACATCTCATT 660
             ||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db    118342 TGGAGGCCAGCGTCCTAG-ATATCTCCCCAGACCCACGAGTCTCTCACATCTCATT 118400

Qy       661 TGGTTCTCGGATTCCAGGCACCCAGAGGGCGGCAGAGAGCCACCCCAGATGCCCCTCTGC 720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    118401 TGGTTCTCGGATTCCAGGCACCCAGAGGGCGGCAGAGAGCCACCCCAGATGCCCCTCTGC 118460

Qy       721 TGATCTTCAGGATTCTTGGAGTCGGGATCTTTCAAGTACCCCCAATGGG 780
             |||||||||||||||||||||||||||||||||||||||||||||||||
Db    118461 TGATCTTCAGGATTCTTGGAGTCAGATCTTTCAAGTTCCCAGGCAGTGG 118520

Qy       781 GAACTCTTTGCAGCTCCTCAGCCTGCGCCCTGGAGCCGTGCAGCCTCAAGCCCAACAG 840
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    118521 GAACTCTTTGCAGCTCCTCAGCCTGCGCCCTGGAGCTCTCAAGCCCCAACCG 118580

Qy       841 CACAGTACCCAAGGGAGGAAGCCAAACCTAAGCGGCGGAAGAAAGTGAGGAGGCCCTTCCA 900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    118581 CACAGTACCCAAGGGAGGAAGCCGAAACCGAAGCGGCGGAAGAAAGTGAGGAGGCCCTTCCA 118640

Qy       901 ACGTTG-ATGCCCCTTCTTTCTTCCTCAAATCAAATGTCAGGGAGTCAAAAGGGCTGT---- 955
             ||||||  |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    118641 ACGTTGAAGGCCCCTTCTTTCTTCCTCAAAACAATGTCAGGGAGTCAGTGGGAAAGGACTGTGTAC 118700

Qy       956 AGCACAGGATGGAGTTGATTATCCCTCCTCCCCAAACACCTAGGAACTGAATCTTTTT 1015
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    118701 AGCGCCGGGATGGAGTTGATTTCCTCCCCAAACACCTAGGAACTGAATCTTTTT 118760

Qy      1016 CTTTTTATTTTTGAGATGATGGAGTCTTGCTCTGTTGCCCAGCTGGAGTGCAGTGGTGTGAT 1075
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    118761 -TTTTTTTTTTTGAGACAGAGTCTCGCTGTCTGTTGCCCAGCTGGGGTGCAGTGGTGTGAT 118819

Qy      1076 CTCAGCTTACTGCAACCTCTGTCTCCCGGGTTCAAGCAATTCTCCCATCTCAGCCTCCTG 1135
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    118820 CTTGGTTACTGCAACCTCTGTCTCCCGGGTTCAAGCAATTCTCCTCGGCCTCCTG 118879

Qy      1136 AGTAGCTGGGAT 1147
             ||||||||||||
Db    118880 AGTAGCTGGGAT 118891
```

RESULT 6
BC016941
LOCUS   BC016941   1818 bp   mRNA   linear   PRI 09-NOV-2001
DEFINITION Homo sapiens, clone MGC:21452 IMAGE:3448846, mRNA, complete cds.
ACCESSION   BC016941
VERSION   BC016941.1 GI:16877383
KEYWORDS   MGC.

```
RESULT 5
AC129837
LOCUS           AC129837              190201 bp    DNA     linear   HTG 11-JUN-2003
DEFINITION      Papio anubis clone rp41-88j2, WORKING DRAFT SEQUENCE.
ACCESSION       AC129837
VERSION         AC129837.16  GI:30725963
KEYWORDS        HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE          Papio anubis (olive baboon)
  ORGANISM      Papio anubis
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
                Cercopithecinae; Papio.
REFERENCE       1  (bases 1 to 190201)
  AUTHORS       Prescott,A., Shaikh,T. and Roe,B.A.
  TITLE         Papio anubis BAC Clone rp41-88j2
  JOURNAL       Unpublished
REFERENCE       2  (bases 1 to 190201)
  AUTHORS       Prescott,A. and Roe,B.A.
  TITLE         Direct Submission
  JOURNAL       Submitted (03-AUG-2002) Department Of Chemistry And Biochemistry,
                The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,
                OK 73019, USA
REFERENCE       3  (bases 1 to 190201)
  AUTHORS       Prescott,A., Shaikh,T. and Roe,B.A.
  TITLE         Direct Submission
  JOURNAL       Submitted (11-JUN-2003) Department Of Chemistry And Biochemistry,
                The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,
                OK 73019, USA
COMMENT         On May 15, 2003 this sequence version replaced gi:30698658.
                ----------------- Genome Center
                Center: Department Of Chemistry And Biochemistry
                The University of Oklahoma
                Center code:UOKNOR
                * NOTE: This is a 'working draft' sequence. It currently
                * consists of 1 contigs. Gaps between the contigs
                * are represented as runs of N. The order of the pieces
                * is believed to be correct as given, however the sizes
                * of the gaps between them are based on estimates that have
                * provided by the submitter.
                * This sequence will be replaced
                * by the finished sequence as soon as it is available and
                * the accession number will be preserved.
                *     1  190201: contig of 190201 bp in length.
FEATURES             Location/Qualifiers
     source          1..190201
                     /organism="Papio anubis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9555"
                     /clone="rp41-88j2"
                     /clone_lib="RPCI - 41 Male (Olive) Baboon BAC Library"
BASE COUNT     46020 a  50800 c  49199 g  44180 t    2 others
ORIGIN
```

Query Match   84.0%;   Score 963.6; DB 2; Length 190201;
Best Local Similarity 92.1%;   Pred. No. 2.9e-300;
Matches 1061; Conservative 0; Mismatches 84; Indels 7; Gaps 4;

```
Qy      1 GCCTTGGGTCAAGCAGCAGAATATTAAATAGGCAGGGGAATGCACCTGTAGCTAGTGGGGCCTA 60
          ||||||||||||||||||||| ||||||||||||||||||||||||||| ||||| |||||||||
Db 117742 GCCTTGGGTCAAGCAGCAGAATACTAAATAGGCAGGGGAATGCACCTGTAGCTACTGGGGCTA 117801

Qy     61 CTGCCAGCCTGAAGAGGAAGTGGCTCGACTCTTGACCATGGCTGCGGGTTCCTGAGGATGA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 117802 CTGCCAGCCTGAAGAGGAAGTGGCTCGACTCTTGACCATGGCTGCGGGTTCCTGAGGATGA 117861

Qy    121 GCTAAACCCTTCCATGTACTGGGGTTGAGGCCACCAGCATCAGATGTTGAACTGAAGAA 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 117862 GTTAAACCCTTCCATGTACTGGGGTTTGAGGCCACCAGCATCAGACGCGTGAACTGAAGAA 117921

Qy    181 GGCCTATAGACAGCTGGCAGTGATGTTCATCCTGACAAAATCATCATCCCGGCTGA 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 117922 GGCCTATAGACAGCTGGCAGTGATGTTCATCCCGACAAAACTCATCACCCCGGCTGA 117981
```

```
Qy    121 GCTAAACCCTTCCATGTACTGGGGTTGAGGCCACAGCATCAGATGTTGAACTGAAGAA 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1415 GCTAAACCCTTCCATGTACTGGGGTTGAGGCCACCAGCATCAGATGTTGAACTGAAGAA 1474

Qy    181 GGCCTATAGACAGCTGGCAGTGATGTTCATCCTGACAAAATCATCATCCCGGCTGA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1475 GGCCTATAGACAGCTGGCAGTGATGTTCATCCTGACAAAATCATCATCCCGGCTGA 1534

Qy    241 GGAGGCCTTCAAGGTTTGCCAGCAGCTTGGGACATTGTCAGCAATGCTGAAAAGCGAAA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1535 GGAGGCCTTCAAGGTTTGCCAGCAGCTTGGGACATTGTCAGCAATGCTGAAAAGCGAAA 1594

Qy    301 GGAGTATGAGATGAAACGAATGGCAGAGAATGAGCTCAGTGCCGGTCAGTAAATGAGTTTCT 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1595 GGAGTATGAGATGAAACGAATGGCAGAGAATGAGCTCAGTGCCGGTCAGTAAATGAGTTTCT 1654

Qy    361 GTCCAAGCTGCAAGATGACCTCAAGGAGGCAATGAATACTATGAGTGTAGCCGATGCCA 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1655 GTCCAAGCTGCAAGATGACCTCAAGGAGGCAATGAATACTATGAGTGTAGCCGATGCCA 1714

Qy    421 AGGAAAGCATAGGAGGTTGAAATGACCGGGAACCTAAGAGTGCCAGATACTGTGCTGA 480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1715 AGGAAAGCATAGGAGGTTGAAATGACCGGGAACCTAAGAGTGCCAGATACTGTGCTGA 1774

Qy    481 GTGTAATAGGCTGCATCCTGCTGAGGAAGAGAGACTTTTGGGCAGAGTCAAGCATGTGGG 540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1775 GTGTAATAGGCTGCATCCTGCTGAGGAAGAGAGACTTTTGGGCAGAGTCAAGCATGTTGGG 1834

Qy    541 CCTCAAGATCACCTACTTTGCACTGATGGATGGAAAGTGTATGACATCACAGAGTGGGC 600
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1835 CCTCAAGATCACCTACTTTGCACTGATGGATGGAAAGTGTATGACATCACAGAGTGGGC 1894

Qy    601 TGGATGCCAGCGTGTAGGTATCTCCCCAGATACCCACAGAGTCCCCTATCACATCTCATT 660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1895 TGGATGCCAGCGTGTAGGTATCTCCCCAGATACCCACAGAGTCCCCTATCACATCTCATT 1954

Qy    661 TGGTTCTCCGGATTCCAGGCACCAGAGGGGCGGCAGAGAGCCACCCAGATGCCCCTTCGC 720
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1955 TGGTTCTCCGGATTCCAGGCACCAGAGGGGCGGCAGAGAGCCACCCAGATGCCCCTTCGC 2014

Qy    721 TGATCTTCAGGATTTCTTGAGTCGGATCTTTCAAGTACCCCCAGGGCAGATGCCCAATGG 780
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2015 TGATCTTCAGGATTTCTTGAGTCGGATCTTTCAAGTACCCCCAGGGCAGATGCCCAATGG 2074

Qy    781 GAACTTCTTTTGCAGCTCCTCAGCCTCCTCAGCCTGCCCCTGGAGCCGCTGCAGCCTCTCAAGCCCAACAG 840
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2075 GAACTTCTTTTGCAGCTCCTCAGCCTCCTCAGCCTGCCCCTGGAGCCGCTGCAGCCTCTCAAGCCCAACAG 2134

Qy    841 CACAGTACCCAAGGGAGAAGCCAAACCTAGGCGGCGGAGAAGTGAGGAGGCCCTTCCA 900
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2135 CACAGTACCCAAGGGAGAAGCCAAACCTAGGCGGCGGAGAAGTGAGGAGGCCCTTCCA 2194

Qy    901 ACGTTGATGCCCCTTCTCTTTCTCTTTCTCTTTCTCAGGGAGTCAAAAGGGCTGTAGCAC 960
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2195 ACGTTGATGCCCCTTCTCTTTCTCTTTCTCTTTCTCAGGGAGTCAAAAGGGCTGTAGCAC 2254

Qy    961 AGGATGGAGTTGATTTATTCCCTCCTCCCCCAACACCTAGGAACTGAATCTTTTTCTTTT 1020
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2255 AGGATGGAGTTGATTTATTCCCTCCTCCCCCAACACCTAGGAACTGAATCTTTTTCTTTT 2314

Qy   1021 TATTTTTGAGATGGAGTCTTGCTCTGTTGCCCAGCTGGAGTGCAGTGGTGTGATCTCAG 1080
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2315 TATTTTTGAGATGGAGTCTTGCTCTGTTGCCCAGCTGGAGTGCAGTGGTGTGATCTCAG 2374

Qy   1081 CTTACTGCAACCTCTGTCTCCCGGGTTCAAGCAATTCTCCCATCTCAGCCTCCTGAGTAG 1140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2375 CTTACTGCAACCTCTGTCTCCCGGGTTCAAGCAATTCTCCCATCTCAGCCTCCTGAGTAG 2434

Qy   1141 CTGGGAT 1147
          |||||||
Db   2435 CTGGGAT 2441
```

```
SOURCE        Homo sapiens (human)
  ORGANISM    Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1  (bases 1 to 4274)
  AUTHORS     Strausberg,R.
  TITLE       Direct Submission
  JOURNAL     Submitted (08-APR-2003) National Institutes of Health, Mammalian
              Gene Collection (MGC), Cancer Genomics Office, National Cancer
              Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
              USA
REMARK        NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT       Contact: MGC help desk
              Email: cgapbs-r@mail.nih.gov
              Tissue Procurement: ATCC
              cDNA Library Preparation: Life Technologies, Inc.
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Sequencing Group at the Stanford Human Genome
              Center, Stanford University School of Medicine, Stanford, CA 94305
              Web site:       http://www-shgc.stanford.edu
              Contact: (Dickson, Mark) mcd@paxil.stanford.edu
              Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
              R. M.

              Clone distribution: MGC clone distribution information can be found
              through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
              Series: IRAK Plate: 22 Row: 1 Column: 8
              This clone was selected for full length sequencing because it
              passed the following selection criteria: GenomeScan gene
              prediction.
FEATURES             Location/Qualifiers
     source          1..4274
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="MGC:17391 IMAGE:3913714"
                     /tissue_type="Uterus, leiomyosarcoma"
                     /clone_lib="NIH_MGC_71"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     CDS             93..2201
                     /codon_start=1
                     /product="Similar to RIKEN cDNA 5730551F12 gene"
                     /protein_id="AAH50271.1"
                     /db_xref="GI:29792139"
                     /translation="MAQKFPGERGLYGAHHSGGASLRTLGPSVDPEIPSFSGLRDSAG
                     TAPNGTRCLTEHSGPKHTQHNPAIHWLDPSHGPPGGPGPPRDAEDPDGSFSSEEESG
                     VDQELSKENETGNQKDGNSFLSIPSACNCQGTPGIPEGPYSEGGNGSSNFCHHCTSP
                     ALGEDELEEEYDDEESLKFPSDFSRVSSGKKPPSRQRHFPTKEDTREGGRRDPRSP
                     GRHRLGRKRSQADKRKGLGLWGAEELCQLGGAGFWWLIELLVLVGEYVETCGHLIYAC
                     RQLKSSDLDLFRVWMGVWTGRLGGWAQVMFQFLSQFYCGVGLFTRFKLLGALLLA
                     LALFGLFLQLGWRFLVGLGDRLGWRDKATWLFSWLDSPALQRCLTLLRDSRPWQRLVR
                     IVQWGWLELPWKQNINRQGNAPVASGRYCQPEEVARLITMAGVPEDELNPFHVLGV
                     EATASDVELKKAYRQLAVMVHPDKNHHPRAEEAFKVLRAAWDIVSNAEKREYEMKRM
                     AENELSRSVNEFLSKLQDDLKEAMNTMMCSRCQGKHRRFEMDREPKSARYCAECNRLH
                     PAEEGDFWAESSMLGLKITYFALMDGKVVDITEWAGCQRVGISPDTHRVPYHISFGSR
                     IPGTRGRQRATPDAPPADLQDFLSRIFQVPPGQMPNGNFFAAPQPAPGAAAASKPNST
                     VPKGEAKPKRKKVVRRPFQR"
BASE COUNT     1065 a    1101 c    1143 g     965 t
ORIGIN

Query Match              100.0%;  Score 1147;  DB 9;  Length 4274;
Best Local Similarity   100.0%;  Pred. No. 0;
Matches 1147;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps    0;

Qy        1 GCCTTGGGTCAAGCAGAGATATTAATAGGCAGGGGAATGCACCTGTAGTAGGCGCTA 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1295 GCCTTGGGTCAAGCAGAGATATTAATAGGCAGGGGAATGCACCTGTAGTAGGCGCTA 1354

Qy       61 CTGCCAGCCTGAAGAGAGAAGTGGCTCGACTCTTGACCATGGCTGCGGGGTTCCTGAGGATGA 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1355 CTGCCAGCCTGAAGAGAGAAGTGGCTCGACTCTTGACCATGGCTGCGGGGTTCCTGAGGATGA 1414
```

Left column (sequence alignment blocks):

```
Qy   241 GGAGGCCTTCAAGGTTTTGGGCGAGCAGCTTGGGACAATGCTGAAAAGCGAAA 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   685 GGAGGCCTTCAAGGTTTTGGGCGAGCAGCTTGGGACAATGCTGAAAAGCGAAA 744

Qy   301 GGAGTATGAGATGAAACGAATGGCAGAGAATGAGCTGAGCCGGTCAGTAAATGAGTTTCT 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   745 GGAGTATGAGATGAAACGAATGGCAGAGAATGAGCTGAGCCGGTCAGTAAATGAGTTTCT 804

Qy   361 GTCCAAGCTGCAAGATGACCTCAAGGAGGCAATGAATACTATGATGTGTAGCCGATGCCA 420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   805 GTCCAAGCTGCAAGATGACCTCAAGGAGGCAATGAATACTATGATGTGTAGCCGATGCCA 864

Qy   421 AGGAAAGCATAGGAGGTTGAAAATGGACCGGGAACCTAAGAGTGCCAGATACTGTGCTGA 480
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   865 AGGAAAGCATAGGAGGTTGAAAATGGACCGGGAACCTAAGAGTGCCAGATACTGTGCTGA 924

Qy   481 GTGTAATAGGCTGCATCCTGCTGAGGAAGGAGACTTTTGGGCAGAGTCAAGCATGTTGGG 540
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   925 GTGTAATAGGCTGCATCCTGCTGAGGAAGGAGACTTTTGGGCAGAGTCAAGCATGTTGGG 984

Qy   541 CCTCAAGATCACCTACTTTGCACTGATGGATGGGAAAGGTGTATGACATCACAGAGTGGGC 600
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   985 CCTCAAGATCACCTACTTTGCACTGATGGATGGGAAAGGTGTATGACATCACAGAGTGGGC 1044

Qy   601 TGGATGCCAGCGTGTAGGTATCTCCCAGATACCCACAGAGTCCCTATCACATCTCATT 660
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1045 TGGATGCCAGCGTGTAGGTATCTCCCAGATACCCACAGAGTCCCTATCACATCTCATT 1104

Qy   661 TGGTTCTCGGATTCCAGGCACCAGAGGGCGGCAGAGCCACCCCCAGAGTGCCCCTCTGC 720
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1105 TGGTTCTCGGATTCCAGGCACCAGAGGGCGGCAGAGCCACCCCCAGAGTGCCCCTCTGC 1164

Qy   721 TGATCTTCAGGATTTCTTGAGTCGGATCTTCAAGTACCCCCAGATGCCCAATGG 780
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1165 TGATCTTCAGGATTTCTTGAGTCGGATCTTCAAGTACCCCCAGATGCCCAATGG 1224

Qy   781 GAACTTCTTTGCAGCTCCTCAGCCTGCCCTGGAGCGCGTCCTCAAGCCCAACAG 840
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1225 GAACTTCTTTGCAGCTCCTCAGCCTGCCCTGGAGCGCGTCCTCAAGCCCAACAG 1284

Qy   841 CACAGTACCCAAGGGAGGAGAAGCCAAACCTAAGCGGCGGAAGAAAGTGAGGAGGCCCTTCCA 900
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1285 CACAGTACCCAAGGGAGGAGAAGCCAAACCTAAGCGGCGGAAGAAAGTGAGGAGGCCCTTCCA 1344

Qy   901 ACGTGATGCCCCTTCTTCTTTCCTCAAATCAATGTCAGGAGTCAAAAGGGCTGTAGCAC 960
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1345 ACGTGATGCCCCTTCTTCTTTCCTCAAATCAATGTCAGGAGTCAAAAGGGCTGTAGCAC 1404

Qy   961 AGGATGGAGTTGATTATCCCTCCTCCCCAACACTAGGAACTGAATCTTTTCTTTT 1020
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1405 AGGATGGAGTTGATTATCCCTCCTCCCCAACACTAGGAACTGAATCTTTTCTTTT 1464

Qy  1021 TATTTTTGAGATGGGAGTCTTGCTCTGTCGCCCAGCTGGAGTGCAGTGGCAGTCTCAG 1080
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1465 TATTTTTTGAGATGGGAGTCTTGCTCTGTCGCCCAGCTGGAGTGCAGTGGCAGTCTCAG 1524

Qy  1081 CTTACTGCAACCTCTGTCTCCCGGGTTCAAGCAATTCTCCCATCTCAGCCCTCCTGAGTAG 1140
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1525 CTTACTGCAACCTCTGTCTCCCGGGTTCAAGCAATTCTCCCATCTCAGCCCTCCTGAGTAG 1584

Qy  1141 CTGGGAT 1147
         |||||||
Db  1585 CTGGGAT 1591
```

```
RESULT 4
BC050271
LOCUS       BC050271                4274 bp    mRNA    linear   PRI 11-APR-2003
DEFINITION  Homo sapiens, Similar to RIKEN cDNA 5730551F12 gene, clone
            MGC:17391 IMAGE:3913714, mRNA, complete cds.
ACCESSION   BC050271
VERSION     BC050271.1  GI:29792138
KEYWORDS    MGC.
```

```
RESULT 3
AK055945
LOCUS       AK055945     2343 bp    mRNA    linear    PRI 01-AUG-2002
DEFINITION  Homo sapiens cDNA FLJ31383 fis, clone NHNPC2000206, highly similar
            to Bovine viral diarrhea virus type 2 strain BVDV2-SD1610c
            polyprotein gene.
ACCESSION   AK055945
VERSION     AK055945.1  GI:16550797
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
```

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
AUTHORS     Tashiro,H., Yamazaki,M., Watanabe,K., Kumagai,A., Itakura,S.,
            Fukuzumi,Y., Fujimori,Y., Komiyama,M., Sugiyama,T., Irie,R.,
            Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y.,
            Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H.,
            Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K.,
            Wagatsuma,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A.,
            Oshima,A., Sugiyama,H., Kawakami,B., Suzuki,Y., Sugano,S.,
            Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.
TITLE       NEDO human cDNA sequencing project
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 2343)
AUTHORS     Isogai,T., Otsuki,T. and Sugiyama,T.
TITLE       Direct Submission
JOURNAL     Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute,
            Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            (E-mail:genomics@hri.co.jp, Tel:81-438-52-3375, Fax:81-438-52-3986)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            Economy, Trade and Industry of Japan; cDNA full insert sequencing:
            Research Association for Biotechnology (RAB) cDNA library
            construction: Helix Research Institute (HRI) (supported by Japan
            Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
            HRI, and Biotechnology Center, National Institute of Technology and
            Evaluation; clone selection for full insert sequencing: RAB and
            HRI.
FEATURES             Location/Qualifiers
     source          1..2343
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="NHNPC2000206"
                     /cell_type="normal neural progenitor cells (NHNP5958)"
                     /clone_lib="NHNPC2"
                     /note="cloning vector: pME18SFL3-primary culture, normal
                     neural progenitor cells"
     CDS             113..1351
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="BAB71050.1"
                     /db_xref="GI:16550798"
                     /translation="MGVWTGRLGGWAQVMFQFLSQGFYCCGVGLFTRFLKLLGALLLA
                     LALFLGFLQLGWRFIVGLGDRLGWRDKAATWLFSWLDSPAIQRCLTLLRDSRPWQRLVR
                     IVQWGWLELPWVKQNINRQGNAPVASGRVCQPEEEVARLLTMAGVPEDELNPFHVLGV
                     EATASDVELKKAYRQLAVMVHPDKNHPRAEEAFKVLRAAWDIVSNAEKRKEYEMKRM
                     AENELSRSVNEFLSKLQDDLKEAMNTMMCSRCQGKHRREMDREPKSARYCAECNRLH
                     PAEEGDFWAESSMJGLKITYFALMDGKVDITEWAGCQRVGISPDFHRVPYHISFGSR
                     IPGTRGRQRATPDAPPADLQDFLSRIFQVPPGQMPNGNFFAAPQPAPGAAAASKPNST
                     VPKGEAKPKRKKVRRPFQR"
BASE COUNT      547 a    527 c    633 g    636 t
ORIGIN
```

```
Query Match              100.0%;   Score 1147;   DB 9;   Length 2343;
Best Local Similarity    100.0%;   Pred. No. 0;
Matches 1147;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

Qy      1 GCCTTGGTCAAGCAGCAGAATATTAATAGGCACGGGAATGCACCTGACTCGTAGTAGGCGGGGTTCCTGAGGATGA   60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    445 GCCTTGGTCAAGCAGCAGAATATTAATAGGCACGGGAATGCACCTGACTCGTAGTAGGCGGGGTTCCTGAGGATGA  504

Qy     61 CTGCCAGCCTGCAAGACGAAGTGGCTCGACTCTTGACCATGGCTCGGGGTTCCTGAGGATGA  120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    505 CTGCCAGCCTGCAAGACGAAGTGGCTCGACTCTTGACCATGGCTCGGGGTTCCTGAGGATGA  564

Qy    121 GCTAAACCCTTCCATGTACTGGGGTTGAGGCCACCAGCATCAGATGTTGAACTGAAGAA  180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    565 GCTAAACCCTTCCATGTACTGGGGTTGAGGCCACCAGCATCAGATGTTGAACTGAAGAA  624

Qy    181 GGCCTATAGACAGCTGGCAGTGATGGTTCATCCTGACAAAAATCATCATCCCCGGCTGA  240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    625 GGCCTATAGACAGCTGGCAGTGATGGTTCATCCTGACAAAAATCATCATCCCCGGCTGA  684

Qy    301 GGAGTATGAGATGAAACGAATGGCAGAGAATGAGCTGAGCCGGTCAGTAAATGAGTTTCT  360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    745 GGAGTATGAGATGAAACGAATGGCAGAGAATGAGCTGAGCCGGTCAGTAAATGAGTTTCT  804

Qy    361 GTCCAAGCTGCAAGATGACCTCAAGGAGGCAATGAATACTATGATGTGTAGCCGATGCCA  420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    805 GTCCAAGCTGCAAGATGACCTCAAGGAGGCAATGAATACTATGATGTGTAGCCGATGCCA  864

Qy    421 AGGAAAGCATAGGAGGTTGAAATGACCGGGAACCTTAAGAGTGCCAGATACTGTGCTGA  480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    865 AGGAAAGCATAGGAGGTTGAAATGACCGGGAACCTTAAGAGTGCCAGATACTGTGCTGA  924

Qy    481 GTGTAATAGGCTGCATCCTGCTGCAGGAAGAGAGACTTTGGGCAGAGTCAAGCATGTTGGG  540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    925 GTGTAATAGGCTGCATCCTGCTGCAGGAAGAGAGACTTTGGGCAGAGTCAAGCATGTTGGG  984

Qy    541 CCTCAAGATCACCTACTTTGCACTGATGATGGAAAGGTGATGACATCACAGAGTGGGC  600
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    985 CCTCAAGATCACCTACTTTGCACTGATGATGGAAAGGTGATGACATCACAGAGTGGGC 1044

Qy    601 TGGATGCCAGCGTGTAGGTATCTCCCAGATACCCCAGAGTCCCTATCACATCTCATT  660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1045 TGGATGCCAGCGTGTAGGTATCTCCCAGATACCCCAGAGTCCCTATCACATCTCATT 1104

Qy    661 TGGTTCTCGGATTCCAGGCACCAGAGGGCGGCGAGAGCCCACCCCAGATGCCCTCCGC  720
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1105 TGGTTCTCGGATTCCAGGCACCAGAGGGCGGCGAGAGCCCACCCCAGATGCCCTCCGC 1164

Qy    721 TGATCTTCAGGATTTCTTGAGTCGGATCTCTTCAAGTACCCCCAGGGCAGATGCCCAATGG  780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1165 TGATCTTCAGGATTTCTTGAGTCGGATCTCTTCAAGTACCCCCAGGGCAGATGCCCAATGG 1224

Qy    781 GAACTTCTTTGCCAGCTCCTCAGCCTCTGCCCCTGGAGCGGCGGAGAAAGTGAGGAGGCCCTTCCA  840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1225 GAACTTCTTTGCCAGCTCCTCAGCCTCTGCCCCTGGAGCGGCGGAGAAAGTGAGGAGGCCCTTCCA 1284

Qy    841 CACAGTACCCAAGGAGAAGAGCCAAACCTAAGCGGCGGCGGAGAAAGTGAGGAGGCCCCTTCCA  900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1285 CACAGTACCCAAGGAGAAGAGCCAAACCTAAGCGGCGGCGGAGAAAGTGAGGAGGCCCCTTCCA 1344

Qy    901 ACGTTGATGCCCCTTCTCTTTCTTCTTTCTCTTTCCTCAAATCAATGTCAGGGAGTCAAAAGGGCTGTAGCAC  960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1345 ACGTTGATGCCCCTTCTCTTTCTTCTTTCTCTTTCCTCAAATCAATGTCAGGGAGTCAAAAGGGCTGTAGCAC 1404

Qy    961 AGGATGGAGTTGATTATCCTCCTCCCCCAACACCTAGGAACTGAATCTTTTCTTTT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1405 AGGATGGAGTTGATTATCCTCCTCCCCCAACACCTAGGAACTGAATCTTTTCTTTT 1464

Qy   1021 TATTTTTGAGATGGAGTCTTGCTCTGTTGCCCAGCTGGAGTGCAGTGGTGTGATCTCAG 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1465 TATTTTTGAGATGGAGTCTTGCTCTGTTGCCCAGCTGGAGTGCAGTGGTGTGATCTCAG 1524

Qy   1081 CTTACTGCAACCCTGTCTCCCCGGGTTCAAGCAATTCTCCCATCTCAGCCTCCTGAGTAG 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1525 CTTACTGCAACCCTGTCTCCCCGGGTTCAAGCAATTCTCCCATCTCAGCCTCCTGAGTAG 1584

Qy   1141 CTGGGAT 1147
          |||||||
Db   1585 CTGGGAT 1591
```

FEATURES
source   Incyte Genomics, Inc. (US)
          Location/Qualifiers
          1..1147
          /organism="Homo sapiens"
          /mol_type="genomic DNA"
          /db_xref="taxon:9606"
          /note="Incyte ID No: 1459372CB1"

BASE COUNT   287 a   278 c   314 g   268 t
ORIGIN

Query Match   100.0%;   Score 1147;   DB 6;   Length 1147;
Best Local Similarity   100.0%;   Pred. No. 0;
Matches 1147;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

```
Qy    1  GCCTTGGGTCAAGCAGAATATTAATAGGCAGGGAATGCACCTGTAGCTAGTGGGCGCTA   60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  GCCTTGGGTCAAGCAGAATATTAATAGGCAGGGAATGCACCTGTAGCTAGTGGGCGCTA   60

Qy   61  CTGCCAGCCTGAAGAGGAAGTGGCTCGACTCTTGACCATGGTCTGGGGTTCCTGAGGATGA  120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  CTGCCAGCCTGAAGAGGAAGTGGCTCGACTCTTGACCATGGTCTGGGGTTCCTGAGGATGA  120

Qy  121  GCTAAACCCTTTCCATGTACTGGGGGTTGAGGCCACAGCAGCATCAGATGTTGAACTGAAGAA  180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  GCTAAACCCTTTCCATGTACTGGGGGTTGAGGCCACAGCAGCATCAGATGTTGAACTGAAGAA  180

Qy  181  GGCCTATAGACAGCTGGCAGTGATGGTTCATCCTGACAAAAATCATCATCCCCGGGCTGA  240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  GGCCTATAGACAGCTGGCAGTGATGGTTCATCCTGACAAAAATCATCATCCCCGGGCTGA  240

Qy  241  GGAGGCCTTCAAGGTTTGCGAGCAGCTTGGACATTGTCAGCAATGCTGAAAAGCGAAA  300
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  GGAGGCCTTCAAGGTTTGCGAGCAGCTTGGACATTGTCAGCAATGCTGAAAAGCGAAA  300

Qy  301  GGAGTATGAGATGAAACGAATGGCCAGAGAATGAGCTGAGCCGGTCAGTAAATGAGTTTCT  360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301  GGAGTATGAGATGAAACGAATGGCCAGAGAATGAGCTGAGCCGGTCAGTAAATGAGTTTCT  360

Qy  361  GTCCAAGCTGCAAGATGACCTCAAGGAGGCAATGAATGAATACTATGATGTGTAGCCGATGCCA  420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361  GTCCAAGCTGCAAGATGACCTCAAGGAGGCAATGAATGAATACTATGATGTGTAGCCGATGCCA  420

Qy  421  AGGAAAGCATAGGAGGTTGAAATGACCGGGAACCTAAGAGTGCCAGATACTGTCTGA  480
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421  AGGAAAGCATAGGAGGTTGAAATGACCGGGAACCTAAGAGTGCCAGATACTGTCTGA  480

Qy  481  GTGTAATAGGCTGCATCCTGCTGTGAGGAAGAGACTTTTGGGCAGAGTCAAGCATGTTGGG  540
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481  GTGTAATAGGCTGCATCCTGCTGTGAGGAAGAGACTTTTGGGCAGAGTCAAGCATGTTGGG  540

Qy  541  CCTCAAGATCACCCTACTTTGCACTGATGGAGAAGGTGTATGACATCACAGAGTGGGC  600
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541  CCTCAAGATCACCCTACTTTGCACTGATGGAGAAGGTGTATGACATCACAGAGTGGGC  600

Qy  601  TGGATGCCAGCGTGTAGGTATCTCCCCAGATACCCCAGATCACATCTCATT  660
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601  TGGATGCCAGCGTGTAGGTATCTCCCCAGATACCCCAGATCACATCTCATT  660

Qy  661  TGGTTCTCGATTCCAGGCACCAGAGGGCGGCAGAGCCACCCCAGATGCCCCCTGC  720
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  661  TGGTTCTCGATTCCAGGCACCAGAGGGCGGCAGAGCCACCCCAGATGCCCCCTGC  720

Qy  721  TGATCTTCAGGATTCTTGAGTCGGATCGTTCAAGTCTTCAAGTACACCCCAGGGCCCAATGG  780
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  721  TGATCTTCAGGATTCTTGAGTCGGATCGTTCAAGTCTTCAAGTACACCCCAGGGCCCAATGG  780

Qy  781  GAACTTCTTTGCAGCTCCTCAGCCTCCGCTGCCCCCGTGCTGCAGCCTCTAAGCCCAACAG  840
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  781  GAACTTCTTTGCAGCTCCTCAGCCTCCGCTGCCCCCGTGCTGCAGCCTCTAAGCCCAACAG  840

Qy  841  CACAGTACCCAAGGAGAGAGCCAAACCTAAGCCGGCGGAAGAAAGTGAGGAGGCCCTTCCA  900
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  841  CACAGTACCCAAGGAGAGAGCCAAACCTAAGCCGGCGGAAGAAAGTGAGGAGGCCCTTCCA  900

Qy  901  ACGTTGATGCGCCCTGCTTCTTCCTCAAATCAATGTCAGGGAGTCAAAAGGGCTGTAGCAC  960
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  901  ACGTTGATGCGCCCTGCTTCTTCCTCAAATCAATGTCAGGGAGTCAAAAGGGCTGTAGCAC  960

Qy  961  AGGATGGAGTTGATTATCCCTCCCCCAACACCTAGGAACTGAATCTTTTCTTT  1020
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  961  AGGATGGAGTTGATTATCCCTCCCCCAACACCTAGGAACTGAATCTTTTCTTT  1020

Qy 1021  TATTTTTTGAGATGGAGCTTGTCTCTTGTCTTGTGCCCAGCTGGAGTGCAGTGGTGTGATCTCAG  1080
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1021  TATTTTTTGAGATGGAGCTTGTCTCTTGTCTTGTGCCCAGCTGGAGTGCAGTGGTGTGATCTCAG  1080

Qy 1081  CTTACTGCAACCTCTGTCTCCCGGGTTCAAGCAATTCTCCCATCTCAGCCTCCTGAGTAG  1140
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1081  CTTACTGCAACCTCTGTCTCCCGGGTTCAAGCAATTCTCCCATCTCAGCCTCCTGAGTAG  1140

Qy 1141  CTGGGAT  1147
         |||||||
Db 1141  CTGGGAT  1147
```

RESULT 2
AX713973
LOCUS   AX713973   2343 bp   DNA   linear   PAT 15-APR-2003
DEFINITION   Sequence 657 from Patent EP1293569.
ACCESSION   AX713973
VERSION   AX713973.1   GI:29888901
KEYWORDS   .
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
AUTHORS   Isogai,T., Sugiyama,T., Otsuki,T., Wakamatsu,A., Sato,H., Ishii,S.,
             Yamamoto,J.I., Isono,Y., Hio,Y., Otsuka,K., Nagai,K., Irie,R.,
             Tamechika,I., Seki,N., Yoshikawa,T., Otsuka,M., Nagahari,K. and
             Masuho,Y.
TITLE   Full-length cDNAs
JOURNAL   Patent: EP 1293569-A 657 19-MAR-2003;
             Helix Research Institute (JP) ; Research Association for
             Biotechnology (JP)
FEATURES   Location/Qualifiers
source   1..2343
             /organism="Homo sapiens"
             /mol_type="genomic DNA"
             /db_xref="taxon:9606"
BASE COUNT   547 a   527 c   633 g   636 t
ORIGIN

Query Match   100.0%;   Score 1147;   DB 6;   Length 2343;
Best Local Similarity   100.0%;   Pred. No. 0;
Matches 1147;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

```
Qy    1  GCCTTGGGTCAAGCAGAATATTAATAGGCACCTGAGCTAGTGGGCGCTA   60
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db  445  GCCTTGGGTCAAGCAGAATATTAATAGGCACCTGAGCTAGTGGGCGCTA  504

Qy   61  CTGCCAGCCTGAAGAGGAAGTGGCTCGACTCTTGACCATGCTGGGGTTCCTGAGGATGA  120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  505  CTGCCAGCCTGAAGAGGAAGTGGCTCGACTCTTGACCATGCTGGGGTTCCTGAGGATGA  564

Qy  121  GCTAAACCCTTTCCATGTACTGGGGGTTGAGGCCACAGCAGCATCAGATGTTGAACTGAAGAA  180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  565  GCTAAACCCTTTCCATGTACTGGGGGTTGAGGCCACAGCAGCATCAGATGTTGAACTGAAGAA  624

Qy  181  GGCCTATAGACAGCTGGCAGTGATGGTTCATCCTGACAAAAATCATCATCCTGA  240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  625  GGCCTATAGACAGCTGGCAGTGATGGTTCATCCTGACAAAAATCATCATCCTGA  684

Qy  241  GGAGGCCTTCAAGGTTTGCGAGCAGCTTGGACATTGTCAGCAATGCTGAAAAGCGAAA  300
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  685  GGAGGCCTTCAAGGTTTGCGAGCAGCTTGGACATTGTCAGCAATGCTGAAAAGCGAAA  744
```

OM nucleic - nucleic search, using sw model

Copyright GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

Run on: December 1, 2003, 07:30:31 ; Search time 4369 Seconds
(without alignments)
10740.065 Million cell updates/sec

Title: US-10-049-742-22
Perfect score: 1147
Sequence: 1 gccttgggtcaagcagaata.........gcctcctgagtagctggat 1147

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :   GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1147	100.0	1147	6	AX081438	AX081438 Sequence
2	1147	100.0	2343	6	AX713973	AX713973 Sequence
3	1147	100.0	2343	9	AK055945	AK055945 Homo sapi
4	1147	100.0	4274	9	BC050271	BC050271 Homo sapi
5	963.6	84.0	190201	2	AC129837	AC129837 Papio anu
6	883.4	77.0	1818	9	BC016941	BC016941 Homo sapi
7	867.2	75.6	953	9	AF351784	AF351784 Homo sapi
8	834	72.7	2402	4	AY027882	AY027882 Bos tauru
9	834	72.7	3159	4	AY027881	AY027881 Bos tauru
10	824	71.8	2692	4	AF308815	AF308815 Bos tauru
11	793.6	69.2	2402	10	BC011146	BC011146 Mus muscu
12	781.8	68.2	2392	10	AF351783	AF351783 Rattus no
13 c	765.8	66.8	263371	2	AC098454	AC098454 Rattus no
14	743	64.8	754	9	AF141342	AF141342 Homo sapi
15	661.2	57.6	247309	2	AC097837	AC097837 Rattus no
16	628.2	54.8	747	6	BD020639	BD020639 Novel gen
17	628.2	54.8	747	6	BD100577	BD100577 Novel gen
18 c	453	39.5	118052	9	AC023055	AC023055 Homo sapi
19	400.8	34.9	1253	14	AF268178	AF268178 Bovine vi
20	399	34.8	1258	14	AF268176	AF268176 Bovine vi
21 c	360.2	31.4	255930	2	AC123227	AC123227 Rattus no
22 c	354.2	30.9	198493	2	AC112674	AC112674 Mus muscu
23 c	354.2	30.9	250277	2	AC122757	AC122757 Mus muscu
24	352.8	30.8	1004	14	AF268171	AF268171 Bovine vi
25	352.8	30.8	1233	14	AF268172	AF268172 Bovine vi
26	337.2	29.4	927	14	BDU43603	U43603 Border dise
27	303.6	26.5	873	14	BDU43602	U43602 Border dise
28	298.6	26.0	194221	10	AC113998	AC113998 Mus muscu
29	297	25.9	717	14	BVD125IN1	Z54331 Bovine vira
30	294.6	25.7	1229	14	AF268180	AF268180 Bovine vi
31	294.6	25.7	13196	14	AB078951	AB078951 Bovine vi
32	294.6	25.7	13203	14	AB078952	AB078952 Bovine vi
33	285.2	24.9	752	14	AF268177	AF268177 Bovine vi
34	279.6	24.4	711	14	BVD125IN2	Z54332 Bovine vira
35	278.6	24.3	852	14	AF268179	AF268179 Bovine vi
36	272.8	23.8	12602	14	AF144617	AF144617 Pestiviru
37	251.4	21.9	12119	6	AX057315	AX057315 Sequence
38	251.4	21.9	12573	14	BVDCG	BVDCG Bovine vira
39	251.4	21.9	12578	14	BVI133738	AJ133738 Bovine vi
40	251.4	21.9	12734	6	AR179057	AR179057 Sequence
41	251.4	21.9	12734	14	AF268278	AF268278 Pestiviru
42	251.4	21.9	14078	6	AR123006	AR123006 Sequence
43	251.4	21.9	14078	6	AR214645	AR214645 Sequence
44	251.4	21.9	14578	6	AR094147	AR094147 Sequence
45	251.2	21.9	1114	14	AF268175	AF268175 Bovine vi

ALIGNMENTS

RESULT 1
AX081438
LOCUS   AX081438   1147 bp   DNA   linear   PAT 27-FEB-2001
DEFINITION Sequence 22 from Patent WO0109178.
ACCESSION   AX081438
VERSION   AX081438.1 GI:13170261
KEYWORDS   .
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
AUTHORS   Yue,H., Bandman,O., Tang,Y.T., Baughn,M.R., Azimzai,Y. and Lu,D.A.
TITLE   Human chaperone proteins
JOURNAL   Patent: WO 0109178-A 22 08-FEB-2001;

Pred. No. is the number of results predicted by chance to have a

Query Match   21.9%; Score 251.4; DB 21; Length 12842;
Best Local Similarity   94.2%; Pred. No. 3.3e-66;
Matches 261; Conservative   0; Mismatches 16; Indels   0; Gaps   0;

```
Qy  404 ATGTGTAGCCGATGCCAAGGAAAGCATAGGAGGTTGAAATGGACCGGGAACCTAAGAGT 463
        ||||| ||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db 4962 ATGTGCAGCCGATGCCAGGGAAAGCATAGGAGGTTGAAATGGACCGGGAACCTAAGAGT 5021

Qy  464 GCCAGATACTGTGCTGAGTGTAATAGGCTGCATCCTGCTGAGGAGGAGACTTTTGGGCA 523
        ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db 5022 GCCAGATACTGTGCTGAGTGTAATAGGCTGCATCCTGCTGAGGAGGTGACTTTTGGGCA 5081

Qy  524 GAGTCAAGCATGTTGGGCCTCAAGATCACCTACTTTGCACTGATGGATGGAAAGGTGTAT 583
        ||||| |||||||||||||||||| |||||||||||| |||||||||||||||||||||
Db 5082 GAGTCGAGCATGTTGGGCCTCAAAATCACCTACTTTGCGCTGATGGATGGAAAGGTGTAT 5141

Qy  584 GACATCACAGAGTGGGCTGGATGCCAGCGTGTAGGTATCTCCCCAGATACCCACCAGAGTC 643
        || ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db 5142 GATATCACAGAGTGGGCTGGATGCCAGCGTGTGGGAATCTCCCCAGATACCCACCAGAGTC 5201

Qy  644 CCCTATCACATCCATTTGGTTCTCGGGATTCCAGGCA 680
        || | ||||||| |||||||||| || || |    ||
Db 5202 CCTTGTCACATCTCATTTGGTTCACGGATGCCTTTCA 5238
```

RESULT 15
AAZ36211
ID   AAZ36211 standard; DNA; 13198 BP.
XX
AC   AAZ36211;
XX
DT   11-FEB-2000 (first entry)
XX
DE   Nucleotide sequence of functional HCV-BVDV chimera from pCBV/p7/IRES-pac.
XX
KW   Pseudorevertant; RNA virus; chimera; BVDV; HCV; replication-competent;
KW   5' nontranslated region; 5'NTR; 3' NTR; pestivirus; antiviral;
KW   bovine viral diarrhea virus; NADL; vaccine; ss.
XX
OS   Chimeric - Hepatitis C virus.
OS   Chimeric - Bovine viral diarrhea virus.
XX
PN   WO9955366-A1.
XX
PD   04-NOV-1999.
XX
PF   23-APR-1999; 99WO-US08850.
XX
PR   24-APR-1998; 98US-0082964.
XX
PA   (UNIW) UNIV WASHINGTON.
XX
PI   Rice CM, Frolov I, McBride MS;
XX
DR   WPI; 2000-013359/01.
XX
PT   Chimeric viral RNA, used in vaccine against BVDV -
XX
PS   Example 3; Fig 24; 108pp; English.
XX
CC   The present sequence represents a functional Hepatitis C virus
CC   (HCV)-bovine viral diarrhea virus (BVDV) chimeric virus of the
CC   invention, expressing a dominant selectable marker conferring resistance
CC   to puromycin. The specification describes chimeric viral RNA comprising
CC   a 5' nontranslated region (5'NTR); an open reading frame (ORF) region;
CC   and a 3' NTR; where at least one of the regions is chimeric and
CC   comprises a nucleotide sequence from a pestivirus in operable linkage
CC   with a heterologous nucleotide sequence, preferably from HCV. The
CC   chimeric viral RNA is replication-competent. The chimeric viral RNA
CC   can be used in a method for identifying compounds having antiviral
CC   activity against HCV. When the pestivirus viral nucleotide sequence is
CC   from bovine viral diarrhea virus (BVDV), the chimeric viral RNA can be

CC   used in a vaccine against BVDV.
XX
SQ   Sequence 13198 BP; 3735 A; 3097 C; 3533 G; 2833 T; 0 other;

Query Match   21.9%; Score 251.4; DB 21; Length 13198;
Best Local Similarity   94.2%; Pred. No. 3.3e-66;
Matches 261; Conservative   0; Mismatches 16; Indels   0; Gaps   0;

```
Qy  404 ATGTGTAGCCGATGCCAAGGAAAGCATAGGAGGTTGAAATGGACCGGGAACCTAAGAGT 463
        ||||| ||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db 4089 ATGTGCAGCCGATGCCAGGGAAAGCATAGGAGGTTGAAATGGACCGGGAACCTAAGAGT 4148

Qy  464 GCCAGATACTGTGCTGAGTGTAATAGGCTGCATCCTGCTGAGGAGGAGACTTTTGGGCA 523
        ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db 4149 GCCAGATACTGTGCTGAGTGTAATAGGCTGCATCCTGCTGAGGAGGTGACTTTTGGGCA 4208

Qy  524 GAGTCAAGCATGTTGGGCCTCAAGATCACCTACTTTGCACTGATGGATGGAAAGGTGTAT 583
        ||||| |||||||||||||||||| |||||||||||| |||||||||||||||||||||
Db 4209 GAGTCGAGCATGTTGGGCCTCAAAATCACCTACTTTGCGCTGATGGATGGAAAGGTGTAT 4268

Qy  584 GACATCACAGAGTGGGCTGGATGCCAGCGTGTAGGTATCTCCCCAGATACCCACCAGAGTC 643
        || ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db 4269 GATATCACAGAGTGGGCTGGATGCCAGCGTGTGGGAATCTCCCCAGATACCCACCAGAGTC 4328

Qy  644 CCCTATCACATCCATTTGGTTCTCGGGATTCCAGGCA 680
        || | ||||||| |||||||||| || || |    ||
Db 4329 CCTTGTCACATCTCATTTGGTTCACGGATGCCTTTCA 4365
```

Search completed: December 1, 2003, 07:37:13
Job time : 384 secs

Left column:

```
Qy  404 ATGTGTAGCCGATGCCAAGGAAGCATAGGAGGTTTGAAATGGACCGGGAACCTAAGAGT 463
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 4994 ATGTGTAGCCGATGCCAAGGAAGCATAGGAGGTTTGAAATGGACCGGGAACCTAAGAGT 5053

Qy  464 GCCAGATACTGTGCTGAGTGTAATAGGCTGCATCCTGCTGAGGAAGAGACTTTTGGGCA 523
        |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db 5054 GCCAGATACTGTGCTGAGTGTAATAGGCTGCATCCTGCTGAGGAAGGTGACTTTTGGGCA 5113

Qy  524 GAGTCAAGCATGTTGGGCCTCAAGATCACCTACTTTGCACTGATGATGGAAAGGTGTAT 583
        ||||| |||||||||||||||||||| |||||||||| |||||||||||||||||||||
Db 5114 GAGTCGAGCATGTTGGGCCTCAAATCACCTACTTTGCGCTGATGATGGAAAGGTGTAT 5173

Qy  584 GACATCACAGAGTGGGCTGAGTGGCTGAGGTATCTCCCAGATACCCACAGAGTC 643
        || |||||||||||||||||| ||||| |||| ||||||||||||||||||||
Db 5174 GATATCACAGAGTGGGCTGGATGGCTGGGAATCTCCCAGATACCCACAGAGTC 5233

Qy  644 CCCTATCACATCCATTGGTTCTCGGATTCCAGGCA 680
        || |||||||| |||||||||| ||||| ||
Db 5234 CCTGTGTCACATCTCATTGGTTCACGGATGCCTTTCA 5270
```

RESULT 13
ABA95615
ID   ABA95615 standard; DNA; 12734 BP.
XX
AC   ABA95615;
XX
DT   21-MAR-2002 (first entry)
XX
DE   Chimeric BVDV/HCV NS3-wt sequence.
XX
KW   Pestivirus; Npro; protease; NS3; screening; ds.
XX
OS   Chimeric - Bovine viral diarrhea virus.
OS   Chimeric - Hepatitis C virus.
XX
PN   US6326137-B1.
XX
PD   04-DEC-2001.
XX
PF   25-JUN-1999; 99US-0344456.
XX
PR   25-JUN-1999; 99US-0344456.
XX
PA   (SCHE) SCHERING CORP.
XX
PI   Hong Z, Lai VCH, Lau JYN;
XX
DR   WPI; 2002-121103/16.
XX
PT   Nucleic acid construct encoding chimeric Hepatitis C Virus (HCV)
PT   pestivirus genome where the Npro protease gene is replaced with NS3
PT   protease gene, useful for in vivo screening of compounds which inhibit
PT   HCV infection -
XX
PS   Example 2; Columns 17-28; 20pp; English.
XX
CC   The present invention relates to a nucleic acid construct encoding a
CC   chimeric Hepatitis C virus (HCV)-pestivirus genome. The construct
CC   comprises a pestivirus genome where a Npro pestivirus protease gene is
CC   replaced with a gene encoding a functional HCV NS3 protease. Furthermore,
CC   each junction site recognised by the Npro protease is replaced with a
CC   junction site recognised by the HCV NS3 protease. The construct is useful
CC   for screening compounds that inhibit HCV in vivo by inhibiting HCV
CC   protease, where screening may be in cell culture or in an animal model.
CC   The present sequence is a chimeric clone of BVDV (bovine viral diarrhea
CC   virus)/HCV NS3-wt, which was used to illustrate the present invention.
XX
SQ   Sequence 12734 BP; 4032 A; 2604 C; 3295 G; 2803 T; 0 other;

Query Match   21.9%; Score 251.4; DB 24; Length 12734;
Best Local Similarity 94.2%; Pred. No. 3.3e-66;
Matches 261; Conservative   0; Mismatches 16; Indels   0; Gaps   0;

Right column:

```
Qy  404 ATGTGTAGCCGATGCCAAGGAAGCATAGGAGGTTTGAAATGGACCGGGAACCTAAGAGT 463
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 5150 ATGTGTAGCCGATGCCAAGGAAGCATAGGAGGTTTGAAATGGACCGGGAACCTAAGAGT 5209

Qy  464 GCCAGATACTGTGCTGAGTGTAATAGGCTGCATCCTGCTGAGGAAGAGACTTTTGGGCA 523
        |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db 5210 GCCAGATACTGTGCTGAGTGTAATAGGCTGCATCCTGCTGAGGAAGGTGACTTTTGGGCA 5269

Qy  524 GAGTCAAGCATGTTGGGCCTCAAGATCACCTACTTTGCACTGATGATGGAAAGGTGTAT 583
        ||||| |||||||||||||||||||| |||||||||| |||||||||||||||||||||
Db 5270 GAGTCGAGCATGTTGGGCCTCAAATCACCTACTTTGCGCTGATGATGGAAAGGTGTAT 5329

Qy  584 GACATCACAGAGTGGGCTGAGTGGCTGAGGTATCTCCCAGATACCCACAGAGTC 643
        || |||||||||||||||||| ||||| |||| ||||||||||||||||||||
Db 5330 GATATCACAGAGTGGGCTGGATGGCTGGGAATCTCCCAGATACCCACAGAGTC 5389

Qy  644 CCCTATCACATCCATTGGTTCTCGGATTCCAGGCA 680
        || |||||||| |||||||||| ||||| ||
Db 5390 CCTGTGTCACATCTCATTGGTTCACGGATGCCTTTCA 5426
```

RESULT 14
AAZ36203
ID   AAZ36203 standard; DNA; 12842 BP.
XX
AC   AAZ36203;
XX
DT   11-FEB-2000 (first entry)
XX
DE   Nucleotide sequence of the prototype HCV-BVDV chimera.
XX
KW   Pseudorevertant; RNA virus; chimera; BVDV; HCV; replication-competent;
KW   5' nontranslated region; 5'NTR; 3' NTR; pestivirus; antiviral;
KW   bovine viral diarrhea virus; NADL; vaccine; ss.
XX
OS   Chimeric - Hepatitis C virus.
OS   Chimeric - Bovine viral diarrhea virus.
XX
PN   WO955366-A1.
XX
PD   04-NOV-1999.
XX
PF   23-APR-1999; 99WO-US08850.
XX
PR   24-APR-1998; 98US-0082964.
XX
PA   (UNIW) UNIV WASHINGTON.
XX
PI   Rice CM, Frolov I, McBride MS;
XX
DR   WPI; 2000-013359/01.
XX
PT   Chimeric viral RNA, used in vaccine against BVDV -
XX
PS   Example 5; Fig 19; 108pp; English.
XX
CC   The present sequence represents the prototype Hepatitis C virus
CC   (HCV)-bovine viral diarrhea virus (BVDV) chimeric virus of the
CC   invention. The sequence contains the adapted HCV 5'NTR from
CC   5'NTR/R.3orig and tandem 3'NTR elements from HCV followed by BVDV.
CC   The specification describes chimeric viral RNA comprising a
CC   5' nontranslated region (5'NTR); an open reading frame (ORF) region;
CC   and a 3' NTR; where at least one of the regions is chimeric and
CC   comprises a nucleotide sequence from a pestivirus in operable linkage
CC   with a heterologous nucleotide sequence, preferably from HCV. The
CC   chimeric viral RNA is replication-competent. The chimeric viral RNA
CC   can be used in a method for identifying compounds having antiviral
CC   activity against HCV. When the pestivirus viral nucleotide sequence is
CC   from bovine viral diarrhea virus (BVDV), the chimeric viral RNA can be
CC   used in a vaccine against BVDV.
XX
SQ   Sequence 12842 BP; 4034 A; 2612 C; 3282 G; 2914 T; 0 other;

RESULT 11

AAC86936
ID   AAC86936 standard; DNA; 12119 BP.
XX
AC   AAC86936;
XX
DT   02-APR-2001 (first entry)
XX
DE   Nucleotide sequence of a chimeric BVDV/HCV virus.
XX
KW   Chimeric virus; bovine viral diarrhoea virus; BVDV; hepatitis C virus;
KW   HCV; vaccine; viral inhibitor; antiviral; ss.
XX
OS   Synthetic.
OS   Bovine viral diarrhoea virus.
OS   Hepatitis C virus.
XX
FH   Key   Location/Qualifiers
FT   CDS   386..11893
FT   /*tag= a
XX
PN   WO20007535Z-A2.
XX
PD   14-DEC-2000.
XX
PF   02-JUN-2000; 2000WO-US15527.
XX
PR   04-JUN-1999; 99US-0137817.
XX
PA   (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Nam J, Bukh J, Emerson SU, Purcell RH;
XX
DR   WPI; 2001-071081/08.
DR   P-PSDB; AAB31167.
XX
PT   New nucleic acid comprising a chimeric bovine viral diarrhoea virus
PT   genome in which the (non-)structural region has been replaced by the
PT   (non-)structural region of a hepatitis C virus (HCV) genome.
PT   hepatitis C virus (HCV) genome useful for treating or preventing HCV
PT   signs and symptoms
XX
PS   Disclosure; Page 62-66; 97pp; English.
XX
CC   The specification describes a nucleic acid comprising a chimeric virus
CC   genome, specifically bovine viral diarrhoea virus (BVDV) genome in which
CC   the (non-)structural region has been replaced by the (non-)structural
CC   region of a hepatitis C virus (HCV) genome. The nucleic acids comprising
CC   the chimeric virus and the chimeric virus are useful for identifying
CC   cell lines capable of supporting the replication of these chimeric
CC   viruses, in screening for neutralizing antibodies to HCV of different
CC   genotypes, in the production of HCV-BVDV virions, for the development
CC   of inactivated or attenuated vaccines to prevent HCV-BVDV in a mammal,
CC   in studying the molecular properties of viral enzyme activity in vitro, and in
CC   identifying inhibitors of viral enzyme activity which would be useful
CC   as antiviral agents. Formulations or compositions comprising the
CC   chimeric virions may be used to treat or prevent the signs and symptoms
CC   of HCV. The present sequence represents a chimeric nucleic acid of the
CC   invention.
XX
SQ   Sequence 12119 BP; 3589 A; 2668 C; 3211 G; 2651 T; 0 other;

Query Match   21.9%; Score 251.4; DB 22; Length 12119;
Best Local Similarity   94.2%;   Pred. No. 3.2e-66;
Matches 261; Conservative   0; Mismatches 16; Indels   0; Gaps   0;

Qy   404 ATGTGTAGCCGATGCCAAGGAAAGCATAGGAGGTTTGAAATGACCGGGAACCTAAGAGT 463
          ||||| |||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db   4535 ATGTGCAGCCGATGCCAGGGAAAGCATAGGAGGTTTGAAATGACCGGGAACCTAAGAGT 4594

Qy   464 GCCAGATACTGTGCTGAGTGTAATAGGCTGCATCCTGCTGAGGAAGGAGACTTTTGGGCA 523
          ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4595 GCCAGATACTGTGCTGAGTGTAATAGGCTGCATCCTGCTGAGGAAGGTGACTTTTGGGCA 4654

Qy   524 GAGTCAAGCATGTTGGGCCTCAAGATCACCTACTTTGCACTGATGATGGAAAGGTGTAT 583
          |||| ||||||| |||||| |||||||||||||||||||| |||||||||||||||||||
Db   4655 GAGTCGAGCATGTTGGGCCTCAAATCACCTACTTTGCCGTGATGGATGGAAAGGTGTAT 4714

Qy   584 GACATCACAGAGTGGGCTGGAGCGGTGAGGTATCTCCCCAGATACCCACAGAGTC 643
          |||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   4715 GATATCACAGAGTGGGCTGGATGCCAGCGTGTGGGAATCTCCCCAGATACCCACAGAGTC 4774

Qy   644 CCCTATCACATCCATTTGGTTCTCGGATTCCAGGCA 680
          || ||||||||||||||||||||||||||||||||||
Db   4775 CCTTGTCACATCCATTCATTTGGTTCACGGATGCCCTTTCA 4811

RESULT 12

AAZ36196
ID   AAZ36196 standard; cDNA; 12578 BP.
XX
AC   AAZ36196;
XX
DT   11-FEB-2000 (first entry)
XX
DE   Nucleotide sequence of infectious BVDV NADL protein.
XX
KW   Pseudorevertant; RNA virus; chimera; BVDV; HCV; replication-competent;
KW   5' nontranslated region; 5'NTR; 3' NTR; pestivirus; antiviral;
KW   bovine viral diarrhea virus; NADL; vaccine; ss.
XX
OS   Bovine viral diarrhea virus.
XX
FH   Key   Location/Qualifiers
FT   CDS   386..12352
FT   /*tag= a
FT   /product= "NADL protein"
XX
PN   WO995366-A1.
XX
PD   04-NOV-1999.
XX
PF   23-APR-1999; 99WO-US08850.
XX
PR   24-APR-1998; 98US-0082964.
XX
PA   (UNIW) UNIV WASHINGTON.
XX
PI   Rice CM, Frolov I, McBride MS;
XX
DR   WPI; 2000-013359/01.
DR   P-PSDB; AAY53616.
XX
PT   Chimeric viral RNA, used in vaccine against BVDV -
XX
PS   Disclosure; Fig 11; 108pp; English.
XX
CC   The present sequence encodes the NADL protein of bovine viral diarrhea
CC   virus (BVDV). The sequence is used in the course of the invention, to
CC   produce chimeric RNA viruses. The specification describes chimeric viral
CC   RNA comprising a 5' nontranslated region (5'NTR); an open reading frame
CC   (ORF) region; and a 3' NTR; where at least one of the regions is chimeric
CC   and comprises a nucleotide sequence from a pestivirus in operable linkage
CC   with a heterologous nucleotide sequence, preferably from HCV. The
CC   chimeric viral RNA is replication-competent. The chimeric viral RNA
CC   can be used in a method for identifying compounds having antiviral
CC   activity against HCV. When the pestivirus viral nucleotide sequence is
CC   from bovine viral diarrhea virus (BVDV), the chimeric viral RNA can be
CC   used in a vaccine against BVDV.
XX
SQ   Sequence 12578 BP; 4049 A; 2527 C; 3233 G; 2769 T; 0 other;

Query Match   21.9%; Score 251.4; DB 21; Length 12578;
Best Local Similarity   94.2%;   Pred. No. 3.2e-66;
Matches 261; Conservative   0; Mismatches 16; Indels   0; Gaps   0;

XX
SQ   Sequence 7453 BP; 1980 A; 1633 C; 1692 G; 2147 T; 1 other;

Query Match   39.5%; Score 453; DB 22; Length 7453;
Best Local Similarity 100.0%; Pred. No. 1.5e-128;
Matches 453; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    695 AGAGCCACCCAGATGCCCCTCCTGCTGTGATCTTCAGGATTTCTTGAGTCGGATCTTTCAA 754
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3315 AGAGCCACCCAGATGCCCCTCCTGCTGTGATCTTCAGGATTTCTTGAGTCGGATCTTTCAA 3374

Qy    755 GTACCCCAGGGCAGATGCCAATGGAACTCTTTGCAGCTCCTCCAGCCTGCCCCTGGA 814
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3375 GTACCCCAGGGCAGATGCCAATGGAACTCTTTGCAGCTCCTCCAGCCTGCCCCTGGA 3434

Qy    815 GCCGCTGCAGCCTCTAGCCCACAGCACAGTACCCAAGGGAGAAGCCAAACCTAAGCGG 874
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3435 GCCGCTGCAGCCTCTAGCCCACAGCACAGTACCCAAGGGAGAAGCCAAACCTAAGCGG 3494

Qy    875 CGGAAGAAAGTGAGGAGGAGGCCCTTCCAACGTGATGCGCCCTTCTCTTTCCTCAAATCAATG 934
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3495 CGGAAGAAAGTGAGGAGGAGGCCCTTCCAACGTGATGCGCCCTTCTCTTTCCTCAAATCAATG 3554

Qy    935 TCAGGGAGTCAAAAGGGCTGTAGCAGGATGGAGTTGATTTATCCCTCCCCCAAC 994
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3555 TCAGGGAGTCAAAAGGGCTGTAGCAGGATGGAGTTGATTTATCCCTCCCCCAAC 3614

Qy    995 ACCTAGGAACTGAATCTTTTCTTTTATTTTTGAGATGGAGTCTTGCTCTGTTGCCCA 1054
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3615 ACCTAGGAACTGAATCTTTTCTTTTATTTTTGAGATGGAGTCTTGCTCTGTTGCCCA 3674

Qy   1055 GCTGGAGTGCAGTGGTGTGATCTCAGCTCACTGCAACCTCTGCTCCCGGGTTCAAGCAA 1114
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3675 GCTGGAGTGCAGTGGTGTGATCTCAGCTCACTGCAACCTCTGCTCCCGGGTTCAAGCAA 3734

Qy   1115 TTCTCCCATCTCAGCCTCCTGAGTAGCTGGGAT 1147
          |||||||||||||||||||||||||||||||||
Db   3735 TTCTCCCATCTCAGCCTCCTGAGTAGCTGGGAT 3767
```

RESULT 10
ABL55114
ID   ABL55114 standard; DNA; 884 BP.
XX
AC   ABL55114;
XX
DT   08-OCT-2002 (first entry)
XX
DE   Human cDNA sequence #2 from clone HNTPB82.
XX
KW   Human; HNTPB82; secreted protein; immunosuppressive; food preservative;
KW   antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;
KW   vasotropic; cerebroprotective; nootropic; neuroprotective; antibacterial;
KW   virucide; fungicide; ophthalmological; vulnerary; gene therapy; ELISA;
KW   radioimmunoassay; enzyme linked immunosorbent assay; autoimmune disease;
KW   hyperproliferative disorder; cardiovascular disorder; angiogenesis;
KW   cerebrovascular disorder; nervous system disorder; ocular disorder;
KW   wound healing; food additive; ss.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   CDS   272..883
FT   /*tag= a
FT   /note= "No stop codon given"
FT   sig_peptide   272..400
FT   /*tag= b
FT   mat_peptide   401..883
FT   /*tag= c
XX
PN   WO2022638-A1.
XX
PD   21-MAR-2002.

XX
PF   17-JAN-2001; 2001WO-US01386.
XX
PR   12-SEP-2000; 2000US-232104P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA, Komatsoulis GA, Baker KP, Birse CE, Soppet DR, Olsen HS;
PI   Moore PA, Wei P, Ebner R, Duan DR, Shi Y, Choi GH, Fiscella M;
PI   Ni J;
XX
DR   WPI; 2002-258041/30.
DR   P-PSDB; ABB77045.
XX
PT   New nucleic acid molecules encoding 22 human secreted proteins for
PT   diagnosing or treating e.g. autoimmune diseases, hyperproliferative
PT   disorders, and cardiovascular disorders, and used as food additives or
PT   preservatives -
XX
PS   Disclosure; Page 466-467; 526pp; English.
XX
CC   The sequence represents a cDNA sequence of the invention, isolated
CC   from human clone ID HNTPB82. The invention relates to novel
CC   isolated nucleic acid molecules encoding 22 human secreted proteins. The
CC   proteins of the invention have immunosuppressive, antiarthritic,
CC   antirheumatic, antiproliferative, cytostatic, cardiant, vasotropic,
CC   cerebroprotective, nootropic, neuroprotective, antibacterial, virucide,
CC   fungicide, ophthalmological, and vulnerary activity. The polynucleotides
CC   may have a use in gene therapy. The polynucleotides and polypeptides
CC   encoded by them are used to prevent, treat or ameliorate a medical
CC   condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs,
CC   chickens or sheep. The polynucleotides and polypeptides are also used in
CC   diagnosing a pathological condition or susceptibility to a pathological
CC   condition. The antibodies to the proteins can also be used in alleviating
CC   symptoms associated with the disorders and in diagnostic immunoassays
CC   e.g. radioimmunoassays or enzyme linked immunosorbent assays (ELISA).
CC   Disorders which are diagnosed or treated include autoimmune diseases,
CC   hyperproliferative disorders, cardiovascular disorders, cerebrovascular
CC   disorders, angiogenesis, nervous system disorders, infections caused by
CC   bacteria, viruses and fungi and ocular disorders. The polypeptides can
CC   also be used to aid wound healing and epithelial cell proliferation. The
CC   polypeptides can also be used as a food additive or preservative.
XX
SQ   Sequence 884 BP; 187 A; 178 C; 291 G; 228 T; 0 other;

Query Match   23.8%; Score 272.8; DB 24; Length 884;
Best Local Similarity 95.6%; Pred. No. 1.9e-73;
Matches 302; Conservative 0; Mismatches 12; Indels 2; Gaps 2;

```
Qy      1 GCCTTGGGTCAAGCAGCAGAATATTAATAGGCAGGGAATGCACCTGTAGCTAGTGGCGCTA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    562 GCCTTGGGTCAAGCAGCAGAATATTAATAGGCAGGGAATGCACCTGTAGCTAGTGGCGCTA 621

Qy     61 CTGCCAGCCTGAAGAGGAAGTGGCTCGACTCTTGACCATGGCTGCAGCTT-GGGACATTGT 120
          |||||||||||||||||||||||||| |||||||||||||||||||||| ||||||||||
Db    622 CTGCCAGCCTGAAGAGGAAGTGGCTGCAGCTGGTTCCTGACCATGGCTGCAGCTT-GGGACATTGT 681

Qy    121 GCTAACCCTTTCCATGTACTGGGGTTGAGCCACAGCCACAGCATCAGATGTTGAACTGAAGAA 180
          |||||||||||||||||||||||| ||||||||||||| ||||||||||||||||||||||
Db    682 GCTAAACCCTTTCCATGTACTGGGGGTTGAGCCACAGCATCAGATGTTGAACTGAAGAA 741

Qy    181 GGCCTATAGACAGCTGCAGTGATGGTTCATCCTGACAAAAATCATCATCCCGGGTCTGA 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    742 GGCCTATAGACAGCTGCAGTGATGGTTCATCCTGACAAAAATCATCATCCCGGGTCTGA 801

Qy    241 GGAGGGCCTTCAAGG-TTTTGCCAGCAGCTT-GGGACATTGTCAGCAATGCTGAAAAGCGA 298
          ||||||||||||||| ||||||||||||||| |||     ||||||||||||||||||
Db    802 GGAGGGCCTTCAAGGTTTTGCCAGCAGCTTGGGGAGCAGCTTGTCAGCAATGCTGAAAAGCG 861

Qy    299 AAGGAGTATGAGATGA 314
          ||||| ||||| ||
Db    862 AAGGGAGTATGGAGA 877
```

```
PR   19-MAY-2000;  2000US-0205515.
PR   07-JUN-2000;  2000US-0209467.
PR   28-JUN-2000;  2000US-0214886.
PR   30-JUN-2000;  2000US-0215135.
PR   07-JUL-2000;  2000US-0216647.
PR   07-JUL-2000;  2000US-0216880.
PR   11-JUL-2000;  2000US-0217487.
PR   11-JUL-2000;  2000US-0217496.
PR   14-JUL-2000;  2000US-0218290.
PR   26-JUL-2000;  2000US-0220963.
PR   26-JUL-2000;  2000US-0220964.
PR   14-AUG-2000;  2000US-0224518.
PR   14-AUG-2000;  2000US-0224519.
PR   14-AUG-2000;  2000US-0225213.
PR   14-AUG-2000;  2000US-0225214.
PR   14-AUG-2000;  2000US-0225266.
PR   14-AUG-2000;  2000US-0225267.
PR   14-AUG-2000;  2000US-0225268.
PR   14-AUG-2000;  2000US-0225270.
PR   14-AUG-2000;  2000US-0225447.
PR   14-AUG-2000;  2000US-0225757.
PR   14-AUG-2000;  2000US-0225758.
PR   14-AUG-2000;  2000US-0225759.
PR   18-AUG-2000;  2000US-0226279.
PR   22-AUG-2000;  2000US-0226681.
PR   22-AUG-2000;  2000US-0226868.
PR   22-AUG-2000;  2000US-0227182.
PR   23-AUG-2000;  2000US-0227009.
PR   30-AUG-2000;  2000US-0228924.
PR   01-SEP-2000;  2000US-0229287.
PR   01-SEP-2000;  2000US-0229343.
PR   01-SEP-2000;  2000US-0229344.
PR   01-SEP-2000;  2000US-0229345.
PR   05-SEP-2000;  2000US-0229509.
PR   05-SEP-2000;  2000US-0229513.
PR   06-SEP-2000;  2000US-0230437.
PR   06-SEP-2000;  2000US-0230438.
PR   08-SEP-2000;  2000US-0231242.
PR   08-SEP-2000;  2000US-0231243.
PR   08-SEP-2000;  2000US-0231244.
PR   08-SEP-2000;  2000US-0231413.
PR   08-SEP-2000;  2000US-0231414.
PR   08-SEP-2000;  2000US-0232080.
PR   08-SEP-2000;  2000US-0232081.
PR   12-SEP-2000;  2000US-0231968.
PR   14-SEP-2000;  2000US-0232397.
PR   14-SEP-2000;  2000US-0232398.
PR   14-SEP-2000;  2000US-0232399.
PR   14-SEP-2000;  2000US-0232400.
PR   14-SEP-2000;  2000US-0232401.
PR   14-SEP-2000;  2000US-0233063.
PR   14-SEP-2000;  2000US-0233064.
PR   14-SEP-2000;  2000US-0233065.
PR   21-SEP-2000;  2000US-0234223.
PR   21-SEP-2000;  2000US-0234274.
PR   25-SEP-2000;  2000US-0234997.
PR   25-SEP-2000;  2000US-0234998.
PR   26-SEP-2000;  2000US-0235484.
PR   27-SEP-2000;  2000US-0235834.
PR   29-SEP-2000;  2000US-0235836.
PR   29-SEP-2000;  2000US-0236327.
PR   29-SEP-2000;  2000US-0236367.
PR   29-SEP-2000;  2000US-0236368.
PR   29-SEP-2000;  2000US-0236369.
PR   29-SEP-2000;  2000US-0236370.
PR   02-OCT-2000;  2000US-0236802.
PR   02-OCT-2000;  2000US-0237037.
PR   02-OCT-2000;  2000US-0237038.
PR   02-OCT-2000;  2000US-0237039.
PR   02-OCT-2000;  2000US-0237040.
PR   13-OCT-2000;  2000US-0239935.
PR   13-OCT-2000;  2000US-0239937.
PR   20-OCT-2000;  2000US-0240960.
PR   20-OCT-2000;  2000US-0241221.
PR   20-OCT-2000;  2000US-0241785.
PR   20-OCT-2000;  2000US-0241786.
PR   20-OCT-2000;  2000US-0241787.
PR   20-OCT-2000;  2000US-0241808.
PR   20-OCT-2000;  2000US-0241809.
PR   20-OCT-2000;  2000US-0241826.
PR   01-NOV-2000;  2000US-0244617.
PR   08-NOV-2000;  2000US-0246474.
PR   08-NOV-2000;  2000US-0246475.
PR   08-NOV-2000;  2000US-0246476.
PR   08-NOV-2000;  2000US-0246477.
PR   08-NOV-2000;  2000US-0246478.
PR   08-NOV-2000;  2000US-0246523.
PR   08-NOV-2000;  2000US-0246524.
PR   08-NOV-2000;  2000US-0246525.
PR   08-NOV-2000;  2000US-0246526.
PR   08-NOV-2000;  2000US-0246527.
PR   08-NOV-2000;  2000US-0246528.
PR   08-NOV-2000;  2000US-0246532.
PR   08-NOV-2000;  2000US-0246609.
PR   08-NOV-2000;  2000US-0246610.
PR   08-NOV-2000;  2000US-0246611.
PR   08-NOV-2000;  2000US-0246613.
PR   17-NOV-2000;  2000US-0249207.
PR   17-NOV-2000;  2000US-0249208.
PR   17-NOV-2000;  2000US-0249209.
PR   17-NOV-2000;  2000US-0249210.
PR   17-NOV-2000;  2000US-0249211.
PR   17-NOV-2000;  2000US-0249212.
PR   17-NOV-2000;  2000US-0249213.
PR   17-NOV-2000;  2000US-0249214.
PR   17-NOV-2000;  2000US-0249215.
PR   17-NOV-2000;  2000US-0249216.
PR   17-NOV-2000;  2000US-0249217.
PR   17-NOV-2000;  2000US-0249218.
PR   17-NOV-2000;  2000US-0249244.
PR   17-NOV-2000;  2000US-0249245.
PR   17-NOV-2000;  2000US-0249264.
PR   17-NOV-2000;  2000US-0249265.
PR   17-NOV-2000;  2000US-0249297.
PR   17-NOV-2000;  2000US-0249299.
PR   17-NOV-2000;  2000US-0249300.
PR   01-DEC-2000;  2000US-0250160.
PR   01-DEC-2000;  2000US-0250391.
PR   05-DEC-2000;  2000US-0251030.
PR   05-DEC-2000;  2000US-0251988.
PR   05-DEC-2000;  2000US-0256719.
PR   06-DEC-2000;  2000US-0251479.
PR   08-DEC-2000;  2000US-0251856.
PR   08-DEC-2000;  2000US-0251868.
PR   08-DEC-2000;  2000US-0251869.
PR   08-DEC-2000;  2000US-0251989.
PR   08-DEC-2000;  2000US-0251990.
PR   11-DEC-2000;  2000US-0254097.
PR   05-JAN-2001;  2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA, Barash SC,  Ruben SM;
XX
DR   WPI; 2001-46570/50.
XX
PT   Isolated nucleic acid molecule encoding a reproductive system antigen
     is used in preventing, treating or ameliorating a medical condition -
XX
PS   Disclosure; SEQ ID NO 8949; 1297pp + Sequence Listing; English.
XX
CC   The present invention provides the protein and coding sequences of a
CC   number of human reproductive system related antigens. These can be used
CC   in the prevention and treatment of reproductive system disorders,
CC   including cancer. The present sequence is a genomic sequence encoding a
CC   protein of the invention.
```

```
CC  and neurological disorders (Alzheimer's disease and Parkinson's disease),
CC  endocrine disorders (Addison's disease), gastrointestinal disorders
CC  (inflammatory disorders), liver disorders (biliary liver cirrhosis),
CC  pancreatic and gall bladder disorders, disorders of the large intestine,
CC  developmental and inherited disorders, diseases at the cellular level,
CC  and wound healing and epithelial cell proliferation. (I) or (II) is
CC  useful to prevent skin aging, for preventing hair loss, to maintain

Query Match     39.5%;  Score 453;  DB 22;  Length 7453;
Best Local Similarity   100.0%;  Pred. No. 1.5e-128;
Matches  453;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    695 AGAGGCACCCCAGATGCCCCTGCTGATCTTCAGGATTCTTCAGGATCGGGATCTTTCAA 754
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3315 AGAGCCACCCCAGATGCCCCTGCTGATCTTCAGGATTTCTTCAGGATCGGGATCTTTCAA 3374

Qy    755 GTACCCCCAGGGCAGATGCCCAATGGGAACTTCTTTGCAGCTCCTCAGCCTGCCCCTGGA 814
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3375 GTACCCCCAGGGCAGATGCCCAATGGGAACTTCTTTGCAGCTCCTCAGCCTGCCCCTGGA 3434

Qy    815 GCCGCTGCAGCCTCTAAGCCAACAGCACAGTACCCAAGGGAGAAGCCAAACCTAAGCGG 874
          ||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db   3435 GCCGCTGCAGCCTCTAAGCCCAACAGCACAGTACCCAAGGGAGAAGCCAAACCTAAGCGG 3494

Qy    875 CGGAAGAAAGTGAGGAGGCCCTTCCAACGTGATGCCCCTTCCTTCTTCCTTCCTCAAATCAATG 934
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3495 CGGAAGAAAGTGAGGAGGCCCTTCCAACGTGATGCCCCTTCCTTCTTCCTTCCTCAAATCAATG 3554

Qy    935 TCAGGGAGTCAAAAGGGCTGTAGCACAGGATGGAGTTTGATTATCCCTCCTCCCCAAC 994
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3555 TCAGGGAGTCAAAAGGGCTGTAGCACAGGATGGAGTTTGATTATCCCTCCTCCCCAAC 3614

Qy    995 ACCTAGGAACTGAATCTTTCTTTTATTTTCTTTATTTTTGAGATGGAGTCTTGCTCTGTTGCCCA 1054
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3615 ACCTAGGAACTGAATCTTTCTTTTATTTTCTTTATTTTTGAGATGGAGTCTTGCTCTGTTGCCCA 3674

Qy   1055 GCTGGAGTGCAGTGGTGTGATCTCAGCTTACTGCAACCTCTGTCTCCCGGGTTCAAGCAA 1114
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3675 GCTGGAGTGCAGTGGTGTGATCTCAGCTTACTGCAACCTCTGTCTCCCGGGTTCAAGCAA 3734

Qy   1115 TTCTCCCATCTCAGCCTCCTGAGTAGCTGGGAT 1147
          |||||||||||||||||||||||||||||||||
Db   3735 TTCTCCCATCTCAGCCTCCTGAGTAGCTGGGAT 3767

RESULT 9
AAL06261
ID   AAL06261 standard; DNA; 7453 BP.
XX
AC   AAL06261;
XX
DT   21-NOV-2001  (first entry)
XX
DE   Human reproductive system related antigen DNA SEQ ID NO: 8949.
XX
KW   Human; reproductive system related antigen; reproductive system disorder;
KW   cancer; gene therapy; ds.
XX
OS   Homo sapiens.
XX
PN   WO20015320-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01339.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA, Barash SC, Ruben SM;
XX
DR   WPI; 2001-476223/51.
XX
PT   Novel isolated prostate gland related polypeptide useful for diagnosis
PT   and treatment of disorders of prostate such as prostatodystonia,
PT   prostatosis, prostatitis, benign prostatic hypertrophy and malacoplakia
PT   -
XX
PS   Claim 1; SEQ ID No 339; 512pp; English.
XX
CC   The invention relates to novel isolated prostate gland related nucleic
CC   acids (I) and polypeptides (II). (I) and (II) are useful for diagnosis,
CC   prognosis, prevention, and/or treatment of diseases and/or disorders of
CC   the prostate such as acute non-bacterial prostatitis, chronic non-
CC   bacterial prostatitis, acute bacterial prostatitis, prostatodystonia,
CC   prostatosis, granulomatous prostatitis, malacoplakia, benign prostatic
CC   hypertrophy or hyperplasia, and prostate neoplastic disorders, including
CC   adenocarcinomas, transitional cell carcinomas, ductal carcinomas, and
CC   squamous cell carcinomas. (I), (II) and antibody to (II) are useful for
CC   diagnosing and treating reproductive system disorders (Paget's disease),
CC   autoimmune disorders (systemic lupus erythematosus, rheumatoid
CC   arthritis), blood-related disorders (sickle cell anaemia),
CC   hyperproliferative disorders, urinary system disorders
CC   (glomerulonephritis), cardiovascular disorders (arrhythmias),
CC   respiratory disorders, musculoskeletal system disorders, neural activity
```

```
Db    397  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
           CTGAGGAAGAGACTTTTGGGCAGAGTCAAGCATGTTGGGCCTCAAGATCACCTACTTG 456

Qy    561  CACTGATGGATGGAAAGGTGTATGACATCACAGAGTGGCTCGATGCCAG-CGTGTAGGT 619
           |||||||||||||||||||||||||||||||||||||||||| |||||||
Db    457  CACTGATGGATGGAAAGGTGTATGACATCACAGAGTGGCTCGATGCCAGCCGTGTAGGT 516

Qy    620  ATCTCCCAGATACCCACAGAGTCCCTATCACATCTCATTGGTTCTCGGATTCCCAGGC 679
           |||||||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db    517  ATCTCCCAGATACCCACAGAGTCCCTATCACATCTCATTTGGTTCTCGGATTCCCAGGC 576

Qy    680  ACCAGAGGGCGGCAGAGAGCCACCCCAGATGCCCCTCCTGCTCATTCAGGATTCTTG 739
           |||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db    577  ACCAGAGGGCGGCAGAGAGCCACCCCAGATGCCCCTCCTGCTCATTCAGGATTCTTG 636

Qy    740  AGTCGGATCTTTCAAGTACCCCCAGGGCAGATGCCCCAATGGGAACTTCTTTGCAGCTCCT 799
           |||| ||||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db    637  AGTCGGAATCTTTCAAGTACCCCCAGGGCCANATGCCCCAATGGGAACTTCTTTGCAGCTTCT 696

Qy    800  CAGCCTGCCC 809
           ||||||||||
Db    697  AACCTTGCCC 706
```

RESULT 8
AAS30481
ID   AAS30481 standard; DNA; 7453 BP.
XX
AC   AAS30481;
XX
DT   21-NOV-2001 (first entry)
XX
DE   DNA encoding novel prostate gland antigen, Seq ID No 339.
XX
KW   Human; nootropic; neuroprotective; cytostatic; antiparkinsonian;
KW   antianaemic; dermatological; immunosuppressive; antiinflammatory;
KW   antiarthritic; antirheumatic; virucide; hepatotropic; nephrotropic;
KW   osteopathic; prostate gland; prostatitis; adenocarcinoma; hair loss;
KW   prostatosis; malacoplakia; adenocarcinoma; benign prostatic hypertrophy;
KW   hyperplasia; carcinoma; prostate neoplastic disorder; skin aging;
KW   reproductive system disorder; autoimmune disorder; urinary system;
KW   systemic lupus erythematosus; rheumatoid arthritis; cardiovascular;
KW   blood-related disorder; hyperproliferative disorder; respiratory;
KW   neurological disorder; endocrine disorder; inflammatory disorder;
KW   liver disorder; wound healing; food preservative; ds.
XX
OS   Homo sapiens.
XX
PN   WO200015447-A1.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01330.
XX

PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189974.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PK   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.

CC inflammations (e.g. bacterial or viral meningitis or encephalitis),
CC testicular disorders (e.g. nontuberculous granulomatous orchitis),
CC connective tissue disorders, or heart disorders (e.g. ischaemic heart
CC disease or atherosclerosis). The compositions and methods may also be
CC used in assessing the histological type of neoplasm associated with
CC ovarian cancer, monitoring the progression of ovarian cancer.
CC determining whether ovarian cancer has metastasized or is likely to
CC metastasize, selecting a composition for inhibiting ovarian cancer,
CC assessing the ovarian carcinogenic potential of a compound, or
CC inhibiting ovarian cancer or at risk of developing ovarian cancer. The
CC present nucleic acid sequence encodes one of the ovarian cancer markers
CC described in the invention.
XX
SQ Sequence 754 BP; 193 A; 185 C; 200 G; 176 T; 0 other;

Query Match   64.8%; Score 743; DB 24; Length 754;
Best Local Similarity 99.9%; Pred. No. 1.4e-218;
Matches   754; Conservative   0; Mismatches   0; Indels   1; Gaps   1;

Qy   301 GGAGTATGAGATGAAACGAATGGCAGAGAATGAGCTGAGCCGGTCAGTAAATGAGTTTCT 360
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GGAGTATGAGATGAAACGAATGGCAGAGAATGAGCTGAGCCGGTCAGTAAATGAGTTTCT 60

Qy   361 GTCCAAGCTGCAAGATGACCTCAAGGAGGCAATGAATGAATACTATGATGTGTAGCCGATGCCA 420
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GTCCAAGCTGCAAGATGACCTCAAGGAGGCAATGAATGAATACTATGATGTGTAGCCGATGCCA 120

Qy   421 AGGAAAGCATAGGAGGTTTGAAATGACCGGGAACCTAAGAGTGCCAGATACTGTGCTGA 480
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 AGGAAAGCATAGGAGGTTTGAAATGACCGGGAACCTAAGAGTGCCAGATACTGTGCTGA 180

Qy   481 GTGTAATAGGCTGCATCCTGCTGAGGAAGGAGACTTTGGGCAGAGTCAAGCATGTTGGG 540
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 GTGTAATAGGCTGCATCCTGCTGAGGAAGGAGACTTTGGGCAGAGTCAAGCATGTTGGG 240

Qy   541 CCTCAAGATCACCTACTTGCACTGATGGATGGAAAGGTGTATGACATCACAGAGTGGGC 600
         ||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db   241 CCTCAAGATCACCTACTTTGCACTGATGGATGGAAAGGTGTATGACATCACAGAGTGGGC 300

Qy   601 TGGGATGCCAGCGTGTAGGTATCTCCCCAGATACCCCACCACAGAGTCCCTATCACATCTCATT 660
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 TGGGATGCCAGCGTGTAGGTATCTCCCCAGATACCCCACCACAGAGTCCCTATCACATCTCATT 360

Qy   661 TGGTTCTCGGATTCCAGGCACCAGAGGGCGGCAGAGAGCCACCCCAGATGCCCCTCCTGC 720
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 TGGTTCTCGGATTCCAGGCACCAGAGGGCGGCAGAGAGCCACCCCAGATGCCCCTCCTGC 420

Qy   721 TGATCTTCAGGATTCTTGAGTCGGATCTTTCAGGATCTTCAGTACCCCCAGGGCAGATGCCCAATGG 780
         |||||||||||||||||||||||||||||||||||||||||| ||||||
Db   421 TGATCTTCAGGATTCTTGAGTCGGATCTTTCAGGATCTG-CCAATGG 479

Qy   781 GAACTTCTTTGCAGCTCCTCAGCCTGCCCCTGGAGCCGTGCAGCCTCAAGCCCAACAG 840
         |||||||||||||||||||||||||||||||| |||||||||||| |||||||||||
Db   480 GAACTTCTTTGCAGCTCCTCAGCCTGCCCCTGCGAGCCGTGCAGCCTCTAAGCCCAACAG 539

Qy   841 CACAGTACCCAAGGAGAGAGCCAAACCTAAGCGGCGGAAGAAAGTGAGGAGGCCCTTCCA 900
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   540 CACAGTACCCAAGGAGAGAGCCAAACCTAAGCGGCGGAAGAAAGTGAGGAGGCCCTTCCA 599

Qy   901 ACGTTGATGCCCCTTCTCTTTCCTCAAATCAATGTCAGGGAGTCAAAAGGGCTGTAGCAC 960
         |||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db   600 ACGTTGATGCCCCTCTCTTTCCTCAAATCAATGTCAGGGAGTCAAAAGGGCTGTAGCAC 659

Qy   961 AGGATGGAGTTGATTTATCCCTCTCCCCCAACACCTAGGAACTGAATCTTTTTTCTTTT 1020
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   660 AGGATGGAGTTGATTTATCCCTCTCCCCCAACACCTAGGAACTGAATCTTTTTTCTTTT 719

Qy 1021 TATTTTTGAGATGAGTCTGCTCTGTTGCCCAG 1055
         ||||||||||||||||||||||||||||||||
Db   720 TATTTTTGAGATGAGTCTGCTCTGTTGCCCAG 754

RESULT 7

AAI96802
ID   AAI96802 standard; cDNA; 747 BP.
XX
AC   AAI96802;
XX
DT   13-NOV-2001 (first entry)
XX
DE   Human neuroblastoma expressed polynucleotide SEQ ID NO 2877.
XX
KW   Human; neuroblastoma; malignancy; cancer; tumour marker; N-myc; TrkA; ss.
XX
OS   Homo sapiens.
XX
PN   WO200166719-A1.
XX
PD   13-SEP-2001.
XX
PF   02-MAR-2001; 2001WO-JP01629.
XX
PR   07-MAR-2000; 2000JP-0159195.
XX
PA   (CHIB-) CHIBA PREFECTURE.
PA   (HISM) HISAMITSU PHARM CO LTD.
XX
PI   Nakagawara A;
XX
DR   WPI; 2001-565584/63.
XX
PT   Nucleic acids originating in gene expressed in human neuroblastoma,
PT   useful as probe or primer in diagnosing prognosis of human
PT   neuroblastoma, malignancy and susceptibility indicator or tumour marker
PT   for anti-cancer agents -
XX
PS   Claim 1; Page 2103; 2979pp; Japanese.
XX
CC   The invention relates to novel genes (AAI93926-AAI97963) expressed in
CC   human neuroblastoma. The nucleic acids are applicable as a probe or
CC   primer in diagnosing the prognosis of human neuroblastoma, malignancy and
CC   susceptibility indicators or tumour markers for anti-cancer agents. The
CC   gene information for diagnosing prognosis is related to factors similar
CC   to that for N-myc and TrkA genes.
XX
SQ   Sequence 747 BP; 191 A; 167 C; 200 G; 165 T; 24 other;

Query Match   54.8%; Score 628.2; DB 22; Length 747;
Best Local Similarity 97.0%; Pred. No. 3.7e-183;
Matches   650; Conservative   0; Mismatches   19; Indels   1; Gaps   1;

Qy   141 TGGGGGTTGAGCCCAGCATCAGATGTGAACTGAAGAAGGCCTATAGACAGCTGGCAG 200
         ||||| || ||||| ||||| |||||||||||||||||||||||||| ||||||
Db   37 TGGGCACTGTTGGCCTACTGGCAGATGTTGAACTGAAGAAGGCCTATAGACACTGGCAG 96

Qy   201 TGATGGTTCATCCTGACAAAAATCATCCCCGGGCCTGAGGGGCCTTCAGGTTTGC 260
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   97 TGATGGTTCATCCTGACAAAAATCATCCCCGGGCCTGAGGGGCCTTCAGGTTTGC 156

Qy   261 GAGCAGCTTGGGACATTGTCAGCAATGTGTGAGATGAGAAGGAA 320
         ||||||||||||||||||||||||||||||||||||||||||||
Db   157 GAGCAGCTTGGGACATTGTCAGCAATGTGTGAGATGAGAAGGAA 216

Qy   321 TGGCAGAGAATGAGCTGAGCCGGTCAGTAAATGAGTTTCGTCGTCCAAGATGACC 380
         |||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db   217 TGGCAGAGAATGAGCTGAGCCGGTCAGTAAATGAGTTTCTGTCGTCCAAGATGACC 276

Qy   381 TCAAGGAGGCAATGAATGATACTATGTAGCCGATGCCAAGGAAAGCATAGGAGGTTG 440
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   277 TCAAGGAGGCAATGAATGATACTATGTAGCCGATGCCAAGGAAAGCATAGGAGGTTG 336

Qy   441 AAATGACCGGGAACCTAAGAGTGCCAGATACTGTGTGAGTGTAATAGGCTGCATCCTG 500
         ||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db   337 AAATGACCGGGAACCTAAGAGTGCCAGATACTGTGCTGCTGAGTGTAATAGGCTGCATCCTG 396

Qy   501 CTGAGGAAGGAGACTTTTGGGCGAGCAGCATGTGGCGCTCAGCATGTTGGGCCTCAAGATCACCTACTTTG 560

CC   immunogen), binding partners of the polypeptides, a method for
CC   identifying an agent useful in modulating mesenchymal cell
CC   differentiation induction activity of a molecule, a method of diagnosing
CC   a condition characterized by aberrant expression of a nucleic acid
CC   molecule or its expression product; a method for determining regression,
CC   progression or onset of cartilaginous tissue degeneration condition in a
CC   subject characterised by aberrant expression of a nucleic acid molecule
CC   or its expression product, a method for treating a cartilaginous tissue
CC   degeneration condition, a method for treating a subject to reduce the
CC   risk of cartilaginous tissue degeneration condition developing in the
CC   subject, a method for identifying a candidate agent for treating a
CC   cartilaginous tissue degeneration condition, and a solid-phase nucleic
CC   acid molecule array consisting essentially of a set of nucleic acid
CC   molecule as cited above (or known from known genes shown to be
CC   differentially expressed in developing mesenchymal cells using the
CC   technique of representational difference analysis, RDA), its expression
CC   products or fragments, fixed to a solid substrate. The nucleic acids,
CC   polypeptides and agents are useful for treating cartilaginous tissue
CC   degeneration conditions such as osteoarthritis, rheumatoid arthritis,
CC   gout arthritis, adjuvant arthritis, arthritis deformans, infectious
CC   arthritis or osterochondrosis. The present sequence is a cDNA from
CC   a known gene differentially expressed in developing mesenchymal cells.
XX
SQ   Sequence 754 BP; 185 A; 193 C; 200 G; 176 T; 0 other;

Query Match   64.8%; Score 743; DB 24; Length 754;
Best Local Similarity 99.9%; Pred. No. 1.4e-218;
Matches 754; Conservative   0; Mismatches   0; Indels   1; Gaps   1;

Qy   301 GGAGTATGAGATGAAACGAATGGCAGAGAATGAGCTGAGCCGGTCAGTAAATGAGTTTCT 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GGAGTATGAGATGAAACGAATGGCAGAGAATGAGCTGAGCCGGTCAGTAAATGAGTTTCT 60

Qy   361 GTCCAAGCTGCAAGATGACCTCAAGGAGGCAATGAATACTATGATGTGTAGCCGATGCCA 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GTCCAAGCTGCAAGATGACCTCAAGGAGGCAATGAATACTATGATGTGTAGCCGATGCCA 120

Qy   421 AGGAAAGCATAGGAGGTTTGAAATGACCGGGAACCTAAGAGTGCCAGATACTGTCTGA 480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 AGGAAAGCATAGGAGGTTTGAAATGACCGGGAACCTAAGAGTGCCAGATACTGTCTGA 180

Qy   481 GTGTAATAGGCTGCATCCTGCTGAGGAAGGAGACTTTTGGGCAGAGTCAAGCATGTTGGG 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 GTGTAATAGGCTGCATCCTGCTGAGGAAGGAGACTTTTGGGCAGAGTCAAGCATGTTGGG 240

Qy   541 CCTCAAGATCACCTACTTTGCACTGATGGATGGAAAGGTGTATGACATCACAGAGTGGGC 600
          ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 CCTCAAGATCACCTACTTTGCACTGATGGATGGAAAGGTGTATGACATCACAGAGTGGGC 300

Qy   601 TGGATGCCAGCGTGTGTAGGTATCTCCCCAGATACCCACAGAGTCCCTATCACATCTCATT 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 TGGATGCCAGCGTGTGTAGGTATCTCCCCAGATACCCACAGAGTCCCTATCACATCTCATT 360

Qy   661 TGGTTCTCGGATTCCAGGCACCAGAGGGGCGCAGAGAGCCACCCCAGATGCCCCTCCTGC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 TGGTTCTCGGATTCCAGGCACCAGAGGGGCGCAGAGAGCCACCCCAGATGCCCCTTCCA 420

Qy   721 TGATCTTCAGGATTCTTGAGTCGGATCTTTCAAGTACCCCCAGGGGCAGCAGATGCCCAATGG 780
          ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db   421 TGATCTTCAGGATTCTTGAGTCGGATCTTTCAAGTACCCCCAGGGGCAGATG-CCAATGG 479

Qy   781 GAACTTCTTTGCAGCTCCAGCTCCTGCCCCTGGAGCCCGCTGCAGCCTCAAGCCCAACAG 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   480 GAACTTCTTTGCAGCTCCAGCTCCTGCCCCTGGAGCCCGCTGCAGCCTCAAGCCCAACAG 539

Qy   841 CACAGTACCCAGGGAGAAGCCAAACCTAAGCCGCGGAAGAAAGTGAGGAGGCCCTTCCA 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   540 CACAGTACCCAGGGAGAAGCCAAACCTAAGCCGCGGAAGAAAGTGAGGAGGCCCTTCCA 599

Qy   901 ACGTTGATGCCCCTTCTTCCTCAAATCAATGTCAGGGAGTCAAAAGGGGTGTAGCAC 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   600 ACGTTGATGCCCCTTCTTCCTCAAATCAATGTCAGGGAGTCAAAAGGGGTCGTAGCAC 659

Qy   961 AGGATGGAGTTTGATTATCCCTCCTCCCCAACACCTAGGAACTGAATCTTTTCTTTT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   660 AGGATGGAGTTTGATTATCCCTCCTCCCCAACACCTAGGAACTGAATCTTTTCTTTT 719

Qy   1021 TATTTTTGAGATGGAGTCTTGCTCTGTTGCCCAG 1055
          ||||||||||||||||||||||||||||||||||
Db   720 TATTTTTGAGATGGAGTCTTGCTCTGTTGCCCAG 754

RESULT 6
ABS76460
ID   ABS76460 standard; cDNA; 754 BP.
XX
AC   ABS76460;
XX
DT   11-DEC-2002 (first entry)
XX
DE   cDNA encoding human ovarian cancer marker OV38.
XX
KW   Human; ovarian cancer; marker; cancer; familial history; brain disorder;
KW   central nervous system disorder; bacterial meningitis; viral meningitis;
KW   Alzheimer's disease; Parkinson's disease; cerebral oedema; hydrocephalus;
KW   brain herniation; inflammation; encephalitis; testicular disorder;
KW   nontuberculous granulomatous orchitis; connective tissue disorder;
KW   heart disorder; ischaemic heart disease; atherosclerosis; neoplasm;
KW   histological type; carcinogenic; ovarian cancer marker; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO20027928-A2.
XX
PD   19-SEP-2002.
XX
PF   14-MAR-2002; 2002WO-US07826.
XX
PR   14-MAR-2001; 2001US-276025P.
PR   14-MAR-2001; 2001US-276026P.
PR   10-AUG-2001; 2001US-311732P.
PR   19-SEP-2001; 2001US-323580P.
PR   19-SEP-2001; 2001US-324967P.
PR   26-SEP-2001; 2001US-325102P.
PR   26-SEP-2001; 2001US-325149P.
XX
PA   (MILL-) MILLENNIUM PHARM INC.
XX
PI   Monahan JE, Gannavarapu M, Hoersch S, Kamatkar S, Kovatis SG;
PI   Meyers RE, Morrisey MP, Olandt PJ, Sen A, Vieby PO, Mills GB;
PI   Bast RC, Lu K, Schmandt RE, Zhao X, Glatt K;
XX
DR   WPI; 2002-723277/78.
DR   P-PSDB; ABG96364.
XX
PT   Assessing whether a patient is afflicted with ovarian cancer, useful in
PT   assessing the stage or progression of the disease, comprises comparing
PT   the expression level of a cancer marker in a sample from a patient and
PT   from a non cancer patient -
XX
PS   Disclosure; Page 300; 481pp; English.
XX
CC   The present invention relates to a new method for assessing whether a
CC   patient is afflicted with ovarian cancer. The method involves comparing
CC   the expression level of a marker in a patient sample and the normal level
CC   of expression of the marker in a control non-ovarian cancer sample, where
CC   the marker is selected from 363 cancer markers described in the
CC   specification. The method of the invention is useful in diagnosing or
CC   characterising cancer, in detecting the presence of cancer as early as
CC   possible, and the recurrence of ovarian cancer. The method may also be of
CC   particular use with patients having an enhanced risk of developing
CC   ovarian cancer (e.g. patients having a familial history of ovarian
CC   cancer). The cancer markers may be used in the management and treatment
CC   of e.g. brain and central nervous system disorders (e.g. bacterial and
CC   viral meningitis, Alzheimer's disease or Parkinson's disease), brain
CC   disorders (e.g. cerebral oedema, hydrocephalus or brain herniations),

```
PD   14-OCT-1999.
XX
PF   29-MAR-1999;   99WO-US06831.
XX
PR   03-APR-1998;   98US-0054956.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Nandabalan K, Kingsmore S;
XX
DR   WPI; 1999-620203/53.
DR   P-PSDB; AAY32126.
XX
PT   Protein complexes, interacting proteins, and related polynucleotides
PT   useful for treating and preventing e.g. atopic, autoimmune or
PT   neurodegenerative diseases -
XX
PS   Claim 21; Fig 7; 172pp; English.
XX
CC   This is the nucleotide sequence of cDNA which codes for a novel
CC   human LYST interacting protein, LIP6 (see AAY32126), that shows
CC   homology to pestivirus Ns2-3. LYST is the human lysosomal
CC   Chediak-Higashi syndrome (CHS) protein. The invention relates to
CC   complexes of LYST or LYST-2 (see AAY32120) with proteins identified
CC   as interacting with LYST or LYST-2 by a modified yeast two-hybrid
CC   assay system. The interacting proteins include 10 novel proteins,
CC   LIP1-10 (see AAY32121-30). Methods of screening the protein complexes
CC   for efficacy in treating and/or preventing atopic diseases (e.g.
CC   asthma, nasal polyps, hay fever rhinitis, urticaria) autoimmune
CC   diseases (e.g. CHS, rheumatoid arthritis, systemic lupus
CC   erythematosus, inflammatory bowel disease, diabetes mellitus,
CC   multiple sclerosis), neurodegenerative disease, certain forms
CC   of cancer, pigmentation disorders, platelet dysfunction and viral
CC   diseases are provided. Nucleic acids (see AAZ34487-96) encoding
CC   LIP1-10, modulation of LIP function by gene therapy, use of
CC   antisense oligonucleotides for suppression of LIP protein
CC   expression, screening for agonists and antagonists, diagnosing or
CC   screening for the presence of a predisposition to a disease or
CC   disorder, and animal models are also disclosed.
XX
SQ   Sequence 754 BP; 193 A; 185 C; 200 G; 176 T; 0 other;

Query Match             64.8%;  Score 743;  DB 20;  Length 754;
Best Local Similarity   99.9%;  Pred. No. 1.4e-218;
Matches  754;  Conservative    0;  Mismatches   0;  Indels    0;  Gaps    1;

Qy      301 GGAGTATGAGATGAAACGAATGGCAGAGAATGAGCTGAGCCGGTCAGTAAATGAGTTTCT 360
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 GGAGTATGAGATGAAACGAATGGCAGAGAATGAGCTGAGCCGGTCAGTAAATGAGTTTCT 60

Qy      361 GTCCAAGCTGCAAGATGACCTCAAGAGGAGGCAATGAATACTATGATGTAGCCGATGCCA 420
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 GTCCAAGCTGCAAGATGACCTCAAGAGGAGGCAATGAATACTATGATGTAGCCGATGCCA 120

Qy      421 AGGAAAGCATAGGAGGTTGAAATGACCCGGAACCTAAGAGTGCCAGATACTGTGCTGA 480
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 AGGAAAGCATAGGAGGTTGAAATGACCCGGAACCTAAGAGTGCCAGATACTGTGCTGA 180

Qy      481 GTGTAATAGGCTGCATCCTGCTGAGGAAGGAAGACTTTGGGCAGAGTCAAGCATGTTGGG 540
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      181 GTGTAATAGGCTGCATCCTGCTGAGGAAGGAAGACTTTGGGCAGAGTCAAGCATGTTGGG 240

Qy      541 CCTCAAGATCACCTACTTTGCACTGATGGATGGAAAGGTGTATGACATCACAGAGTGGGC 600
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      241 CCTCAAGATCACCTACTTTGCACTGATGGATGGAAAGGTGTATGACATCACAGAGTGGGC 300

Qy      601 TGGATGCCAGCGTGTAGGTATCTCCCACGATACCCACAGAGTCCCTATCACATCTCATT 660
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      301 TGGATGCCAGCGTGTAGGTATCTCCCACGATACCCACAGAGTCCCTATCACATCTCATT 360

Qy      661 TGGTTCTCGGATTCCAGGCACCAGAGGGCGCAGAGGCCACCCCCTCCTGC 720
                ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      361 TGGTTCTCGGATTCCAGGCACCAGAGGGCGCAGAGGCCACCCCCTCCTGC 420

Qy      721 TGATCTTCAGGATTCTTCTGAGTCGGATCTTTCAAGTACCCCCAGGGCGAGATGCCCAATGG 780
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db      421 TGATCTTCAGGATTCTTCTGAGTCGGATCTTTCAAGTACCCCCAGGGGCGAGATG-CCAATGG 479

Qy      781 GAACTTCTTTGCAGCTCCTCAGCTGCCCCTGAGCCGCCTGCAGCCTCAAGCCAACAG 840
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      480 GAACTTCTTTGCAGCTCCTCAGCTGCCCCTGAGCCGCCTGCAGCCTCAAGCCAACAG 539

Qy      841 CACAGTACCCAAGGAGGAAGCCAAACTAAGCGGCGAAGAAGTGAGGAGGCCCTTCCA 900
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      540 CACAGTACCCAAGGAGGAAGCCAAACTAAGCGGCGAAGAAGTGAGGAGGCCCTTCCA 599

Qy      901 ACGTTGATGCCCCTTCTTTCTCCTCAAATCAATGTCAGGGAGTCAAAAGGGCTGTAGCAC 960
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      600 ACGTTGATGCCCCTTCTTTCTCCTCAAATCAATGTCAGGGAGTCAAAAGGGCTGTAGCAC 659

Qy      961 AGGATGAGTTGATTTATCCCTCCTCCCCAACACTAGGAACTGAATCTTTTTCTTT 1020
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      660 AGGATGAGTTGATTTATCCCTCCTCCCCAACACTAGGAACTGAATCTTTTTCTTTT 719

Qy     1021 TATTTTTGAGATGGAGTCTTGCTCTGTTGCCCAG 1055
                |||||||||||||||||||||||||||||||||||
Db      720 TATTTTTGAGATGGAGTCTTGCTCTGTTGCCCAG 754


RESULT 5
ABX04184
ID   ABX04184 standard; cDNA; 754 BP.
XX
AC   ABX04184;
XX
DT   10-JAN-2003  (first entry)
XX
DE   Human mRNA differentially expressed in mesenchymal cells #31.
XX
KW   Human; ss; gene; skeletal growth; cartilage degeneration disorder;
KW   chondroblastic phenotype; mesenchymal cell; cartilage formation;
KW   bone formation; arthritis; osteoarthritis; rheumatoid arthritis;
KW   gout arthritis; adjuvant arthritis; arthritis deformans; antigout;
KW   infectious arthritis; osteochondrosis; RDA; antiarthritic; osteopathic;
KW   antirheumatic; antiinflammatory; representational difference analysis.
XX
OS   Homo sapiens.
XX
PN   WO20071927-A2.
XX
PD   19-SEP-2002.
XX
PF   12-MAR-2002; 2002WO-US07787.
XX
PR   12-MAR-2001; 2001US-274980P.
XX
PA   (BGHM ) BRIGHAM & WOMENS HOSPITAL INC.
XX
PI   Yates KE, Mizuno S,  Glowacki J;
XX
DR   WPI; 2002-723276/78.
XX
PT   New nucleic acid molecules capable of promoting chondrogenesis, useful
PT   for diagnosing and treating cartilaginous tissue degeneration
PT   conditions, e.g. osteoarthritis, rheumatoid arthritis, gout arthritis,
PT   or osterochondrosis -
XX
PS   Claim 33; Page 129; 153pp; English.
XX
CC   The invention relates to new isolated nucleic acid molecule comprising a
CC   nucleic acid molecule consisting of a gene differentially expressed
CC   in cells undergoing differentiation from mesenchymal cell to a
CC   chondroblastic phenotype, or hybridising under stringent conditions
CC   to them (or their fragments). Also included are expression vectors,
CC   transformed host cells, expressed polypeptides or peptide fragments
CC   (which induce differentiation of a mesenchymal cell and may be used as an
```

```
OS   Homo sapiens.
XX
FH   Key            Location/Qualifiers
FT   CDS            122..2230
FT                  /tag= a
FT                  /product= "P125-77.22 polypeptide"
XX
PN   WO20022810-A1.
XX
PD   04-APR-2002.
XX
PF   10-SEP-2001; 2001WO-CN01354.
XX
PR   12-SEP-2000; 2000CN-0125190.
XX
PA   (SHAN-) SHANGHAI BIOWINDOW GENE DEV INC.
XX
PI   Mao Y, Xie Y;
XX
DR   WPI; 2002-281319/32.
DR   P-PSDB; ABB77732.
XX
PT   Protein P125-77.22 and encoding polynucleotide, used in diagnosis and
PT   treatment of human mucosal disease caused by BVDV infection -
XX
PS   Claim 6; Page 27-29; 33pp; Chinese.
XX
CC   The present sequence encodes human P125-77.22 polypeptide. The
CC   polypeptide and polynucleotide are used in diagnosis and treatment
CC   of human mucosal disease caused by viral BVDV (undefined) infection.
CC   The polynucleotide may also be used for gene therapy.
XX
SQ   Sequence 3286 BP; 787 A; 785 C; 895 G; 819 T; 0 other;

Query Match              99.6%;  Score 1142.2;  DB 24;  Length 3286;
Best Local Similarity    99.7%;  Pred. No. 0;
Matches 1144;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 GCCTTGGGTCAAGCAGCAGAATATTAATAGGCAGGGAATGCACCTGTAGCTAGTGGGCGCTA 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1324 GCCTTGGGTCAAGCAGCAGAATATTAATAGGCAGGGAATGCACCTGTAGCTAGTGGGCGCTA 1383

Qy     61 CTGCCAGCCTGAAGAGGAAGTGGCTCGACTCTTGACCATGGCCATGGCTCGGGGTTCCTGAGGATGA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1384 CTGCCAGCCTGAAGAGGAAGTGGCTCGACTCTTGACCATGGCCATGGCTCGGGGTTCCTGAGGATGA 1443

Qy    121 GCTAAACCCTTTCCATGTACTGGGGGTTGAGGCCACAGCATCAGATGTTGAACTGAAGAA 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1444 GCTAAACCCTTTCCATGTACTGGGGGTTGAGGCCACAGCATCAGATGTTGAACTGAAGGAA 1503

Qy    181 GGCCTATAGACAGCTGGCAGTGATGGTTCATCCTGACAAAAATCATCATCCCCGGCTGA 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1504 GGCCTATAGACAGCTGGCAGTGATGGTTCATCCTGACAAAAATCATCATCCCCGGCTGA 1563

Qy    241 GGAGGCCTTCAAGGTTTGCGAGCAGCTTGGGACATTGTCAGCAATGCTGAAAAGCGAAA 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1564 GGAGGCCTTCAAGGTTTGCGAGCAGCTTGGGACATTGTCAGCAATGCTGAAAAGCGAAA 1623

Qy    301 GGAGTGATGAGAAACGAATGGCAGAGATGGCTCAGTAAATGAGTTTCT 360
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db   1624 GGAGTATGAGAAACGAATGGCAGAGATGGCTCAGTAAATGAGTTTCT 1683

Qy    361 GTCCAAGCTGCAAGATGACCTCAAGGAGGCAATGAATACTATGATGTGTAGCCGATGCCA 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1684 GTCCAAGCTGCAAGATGACCTCAAGGAGGCAATGAATACTATGATGTGTAGCCGATGCCA 1743

Qy    421 AGGAAAGCATAGGAGGTTTGAAATGGACCGGGAACCTAAGAGTGCCAGATACTGTGCTGA 480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1744 AGGAAAGCATAGGAGGTTTGAAATGGACCGGGAACCTAAGAGTGCTAGATACTGTGCTGA 1803

Qy    481 GTGTAATAGGCTGCATCCTGCTGAGGAAGGAGACTTTGGGCAGAGTCAAGCATGTTGGG 540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1804 GTGTAATAGGCTGCATCCTGCTGAGGAAGGAGACTTTGGGCAGAGTCAAGCATGTTGGG 1863
```

```
Qy    541 CCTCAAGATCACCTACTTTGCACTGATGATGGAAAGGTGTATGACATCACAGAGTGGGC 600
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1864 CCTCAAGATCACCTACTTTGCACTGATGATGGAAAGGTGTATGACATCACAGAGTGGGC 1923

Qy    601 TGGATGCCAGCGTGTAGGTATCTCCCAGATACCCACAGAGTCCCTATCACATCTCATT 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1924 TGGATGCCAGCGTGTAGGTATCTCCCAGATACCCACAGAGTCCCTATCACATCTCATT 1983

Qy    661 TGGTTCTCGGATTCCAGGACCAGAGAGCGGCAGAGAGGCCACCCAGATGCCCCTCTGC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1984 TGGTTCTCGGATTCCAGGACCAGAGAGCGGCAGAGAGCCACCCAGATGCCCCTCTGC 2043

Qy    721 TGATCTTCAGGATTCTTGAGTCGGATCTTCAAGTACCCCCAGGGCAGATGCCCAATGG 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2044 TGATCTTCAGGATTCTTGAGTCGGATCTTCAAGTACCCCCAGGGCAGATGCCCAATGG 2103

Qy    781 GAACTCTTTGCAGCTTCCTCAGCCTGCCCCTGGAGCCGCTGCAGCCTCTAAGCCCAACAG 840
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2104 GAACTTCTTTGCAGCTTCCTCAGCCTGCCCCTGGAGCCGCTGCAGCCTCTAAGCCCAACAG 2163

Qy    841 CACAGTACCCAAGGGAGAAGCCAAACCTAAGGCGGCGGAAGAAGTGAGGAGGCCCTTCCA 900
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2164 CACAGTACCCAAGGGAGAAGCCAAACCTAAGGCGGCGGAAGAAGTGAGGAGGCCCTTCCA 2223

Qy    901 ACGTTGATGCCCCTTCTTTCCTCAAATCATGTCAGGGAGTCAAAAGGGCTGTAGCAC 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2224 ACGTTGATGCCCCTTCTTTCCTCAAATCAATGTCAGGGAGTCAAAAGGGCTGTAGCAC 2283

Qy    961 AGGATGGAGTTGATTATCCTCCCCCAACACCTAGGAACTGAATCTTTTCTTTT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2284 AGGATGGAGTTGATTATCCTCCCCCAACACCTAGGAACTGAATCTTTTCTTTT 2343

Qy   1021 TATTTTTTGAGATGGAGTCTTGCTCTGTTGCCCAGTGGCAGTGGTGTGATCTCAG 1080
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2344 TATTTTTTGAGATGGAGTCTTGCTCTGTTGCCCAGTGGCAGTGGTGTGATCTCAG 2403

Qy   1081 CTTACTGCAACCTCTGTCTCCCGGGTTCAAGCGAATTCTCCCATCTCAGCCTCCTGAGTAG 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2404 CTTACTGCAACCTCTGTCTCCCGGGTTCAAGCGAATTCTCCCATCTCAGCCTCCTGAGTAG 2463

Qy   1141 CTGGGAT 1147
          |||||||
Db   2464 CTGGGAT 2470
```

```
RESULT 4
AA234492
ID   AA234492 standard; cDNA; 754 BP.
XX
AC   AA234492;
XX
DT   01-FEB-2000  (first entry)
XX
DE   Human LYST interacting protein LIP6 cDNA.
XX
KW   LIP6; human; LYST-2; LYST interacting protein; lysosome; CHS;
KW   Chediak-Higashi syndrome; autoimmune disease; rheumatoid arthritis;
KW   systemic lupus erythematosus; inflammatory bowel disease;
KW   diabetes mellitus; multiple sclerosis; atopic disease; asthma;
KW   hay fever; rhinitis; urticaria; nasal polyp; cancer;
KW   neurodegenerative disease; pigmentation disorder; viral disease;
KW   platelet dysfunction; ss.
XX
OS   Homo sapiens.
XX
FH   Key            Location/Qualifiers
FT   CDS            11..586
FT                  /tag= a
FT                  /partial
XX
PN   WO9951741-A2.
XX
```

```
PN   WO20022638-A1.
XX
PD   21-MAR-2002.
XX
PF   17-JAN-2001; 2001WO-US01386.
XX
PR   12-SEP-2000; 2000US-232104P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA, Komatsoulis GA, Baker KP,  Birse CE, Soppet DR, Olsen HS;
PI   Moore PA, Wei P, Ebner R, Duan DR, Shi Y, Choi GH, Fiscella M;
PI   Ni J;
XX
DR   WPI: 2002-258041/30.
DR   P-PSDB; ABB77019.
XX
PT   New nucleic acid molecules encoding 22 human secreted proteins for
PT   diagnosing or treating e.g. autoimmune diseases, hyperproliferative
PT   disorders, and cardiovascular disorders, and used as food additives or
PT   preservatives -
XX
PS   Disclosure; Page 451-452; 526pp; English.
XX
CC   The sequence represents a cDNA sequence of the invention, isolated
CC   from human clone ID HNTPB82. The invention relates to novel
CC   isolated nucleic acid molecules encoding 22 human secreted proteins. The
CC   proteins of the invention have immunosuppressive, antiarthritic,
CC   antirheumatic, antiproliferative, cytostatic, cardiant, vasotropic,
CC   cerebroprotective, nootropic, neuroprotective, antibacterial, virucide,
CC   fungicide, ophthalmological, and vulnerary activity. The polynucleotides
CC   may have a use in gene therapy. The polynucleotides and polypeptides
CC   encoded by them are used to prevent, treat or ameliorate a medical
CC   condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs,
CC   chickens or sheep. The polynucleotides and polypeptides are also used in
CC   diagnosing a pathological condition or susceptibility to a pathological
CC   condition. The antibodies to the proteins can also be used in alleviating
CC   symptoms associated with the disorders and in diagnostic immunoassays
CC   e.g. radioimmunoassays or enzyme linked immunosorbent assays (ELISA).
CC   Disorders which are diagnosed or treated include autoimmune diseases,
CC   hyperproliferative disorders, cardiovascular disorders, cerebrovascular
CC   disorders, angiogenesis, nervous system disorders, infections caused by
CC   bacteria, viruses and fungi and ocular disorders. The polypeptides can
CC   also be used to aid wound healing and epithelial cell proliferation. The
CC   polypeptides can also be used as a food additive or preservative.
XX
SQ   Sequence 2279 BP; 540 A; 518 C; 607 G; 614 T; 0 other;

Query Match             99.9%;  Score 1145.4;  DB 24;  Length 2279;
Best Local Similarity   99.9%;  Pred. No. 0;
Matches 1146;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    0;

Qy          1 GCCTTGGGTCAAGCAGAATATTAATAGGCAGGGGAATGCACCTGTAGTGGGCGCTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        368 GCCTTGGGTCAAGCAGAATATTAATAGGCAGGGGAATGCACCTGTAGTGGGCGCTA 427

Qy         61 CTGCCAGCCTGAAGAGGAAGTGGCTCGACTCTTGACCATGGCTGGGGGTTCCTGAGGATGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        428 CTGCCAGCCTGAAGAGGAAGTGGCTCGACTCTTGACCATGGCTGGGGGTTCCTGAGGATGA 487

Qy        121 GCTAAACCCTTTCCATGCATGGGGGTTGAGGCCCACAGCATCAGATGTTGAACTGAAGAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        488 GCTAAACCCTTTCCATGCATGGGGGTTGAGGCCCACAGCATCAGATGTTGAACTGAAGAA 547

Qy        181 GGCTATAGACAGCTGGCAGTGATGGTTCATCCTGACAAAAATCATCCCCGGGCTGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        548 GGCTATAGACAGCTGGCAGTGATGGTTCATCCTGACAAAAATCATCCCCGGGCTGA 607

Qy        241 GGAGGCCTTCAAGGTTTTGCGAGCAGCTTGGGACATTGTCAGCAATGCTGAAAAGCGAAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        608 GGAGGCCTTCAAGGTTTTGCGAGCAGCTTGGGACATTGTCAGCAATGCTGAAAAGCGAAA 667

Qy        301 GGAGTATGAGATGAAAACGAATGGCAGAGAATGAGCTGAGCCGGTCAGTAAATGAGTTTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        668 GGAGTATGAGATGAAAACGAATGGCAGAGAATGAGCTGAGCCGGTCAGTAAATGAGTTTCT 727

Qy        361 GTCCAAGCTGCAAGATGACCTCAAGGAGGCAATGAATACTATGATGTGTAGCCGATGCCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        728 GTCCAAGCTGCAAGATGACCTCAAGGAGGCAATGAATACTATGATGTGTAGCCGATGCCA 787

Qy        421 AGGAAAGCATAGGAGGTTTGAAATGGACCGGGAACCTAAGAGTGCCAGATACTGTCTGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        788 AGGAAAGCATAGGAGGTTTGAAATGGACCGGGAACCTAAGAGTGCCAGATACTGTCTGCTGA 847

Qy        481 GTGTAATAGGCTGCATCCTGCTGAGGAAGGAGACTTTTGGGCAGAGTCAAGCATGTTGGG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        848 GTGTAATAGGCTGCATCCTGCTGAGGAAGGAGACTTTTGGGCAGAGTCAAGCATGTTGGG 907

Qy        541 CCTCAAGATCACCTACTTTGCACTGATGGATGGAAAGGTGTATGACATCACACAGTGGGC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        908 CCTCAAGATCACCTACTTTGCACTGATGGATGGAAAGGTGTATGACATCACACAGTGGGC 967

Qy        601 TGGATGCCAGCGTGTAGGTATCTCCCAGATACCCACAGAGTCCCTATCACATCTCATT 660
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        968 TGGATGCCAGCGTGTAGGTATCTCCCAGATACCCACAGAGTCCCTATCACATCTCATT 1027

Qy        661 TGGTTCTCGGATTCCAGGCACCAGAGGGCGCAGAGAGCCACCCCAGATGCCCCTCCTGC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1028 TGGTTCTCGGATTCCAGGCACCAGAGGGCGCAGAGAGCCACCCCAGATGCCCCTCCTGC 1087

Qy        721 TGATCTTCAGGATTCTTGAGTGGGATCTTCAAGTACCCCAGGGCAGATGCCCAATGG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1088 TGATCTTCAGGATTCTTGAGTGGGATCTTCAAGTACCCCAGGGCAGATGCCCAATGG 1147

Qy        781 GAACTTCTTGCAGCCTCCAGCCTCGCCCTGCCCTGAGCGCCTGAAGCCAACAG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1148 GAACTTCTTGCAGCCTCCAGCCTCGCCCTGCCCTGAGCGCCTGCCTCTAAGCCCAACAG 1207

Qy        841 CACAGTACCCAAGGAGAAGCCAAACCTAAGCGGCGGAAGAAAGTGAGGAGGCCCCTCCA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1208 CACAGTACCCAAGGAGAAGCCAAACCTAAGCGGCGGAAGAAAGTGAGGAGGCCCCTTCA 1267

Qy        901 ACGTTGATGCCCCTTCTCTTCCTCCAAATCAATGTCAGGGAGTCAAAAGGGGTCTAGCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1268 ACGTTGATGCCCCTTCTCTTCTTCCTCCAAATCAATGTCAGGGAGTCAAAAGGGGTGTAGCAC 1327

Qy        961 AGGATGGAGTTTGATTTATCCCCTCCCCCAACACCTAGGAACTGAATCTTTTCTTTT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1328 AGGATGGAGTTTGATTTATCCCCTCCTCCCCCAACACCTAGGAACTGAATCTTTTCTTTT 1387

Qy       1021 TATTTTTTGAGATGGAGTCTTGCTCTGTTGCCCAGCTGGAGTGCAGTGGTGTGATCTCAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1388 TATTTTTTGAGATGGAGTCTTGCTCTGTCGTTGCCCAGCTGGAGTGCAGTGGTGTGATCTCAG 1447

Qy       1081 CTTACTGCAACCTCTGTCTCCCGGGTTCAGCAATTCTCCCATCTCAGCCTCCTGAGTAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1448 CTTACTGCAACCTCTGTCTCCCGGGTTCAGCAATTCTCCCATCTCAGCCTCCTGAGTAG 1507

Qy       1141 CTGGGAT 1147
              |||||||
Db       1508 CTGGGAT 1514

RESULT 3
ABL56700
ID   ABL56700 standard; cDNA; 3286 BP.
XX
AC   ABL56700;
XX
DT   30-JUL-2002  (first entry)
XX
DE   Nucleotide sequence of human P125-77.22 polypeptide.
XX
KW   Human; P125-77.22; mucosal disease; BVDV infection; gene therapy;
KW   gene; ss.
XX
```

PR   03-AUG-1999;   99US-0146908.
PR   22-OCT-1999;   99US-0160924.
XX
PA   (INCY-) INCYTE GENOMICS INC.
PI   Yue H, Bandman O, Tang YT, Baughn MR, Azimzai Y, Lu DAM;
XX
DR   WPI; 2001-159853/16.
DR   P-PSDB; AAB67455.
XX
PT   New human chaperone proteins and polynucleotides, useful in diagnosing,
PT   treating and preventing reproductive, eye, neuromuscular, metabolic,
PT   autoimmune or inflammatory disorders -
XX
PS   Claim 5; Page 95-96; English.
XX
CC   The present sequence encodes a human chaperone polypeptide. Human
CC   chaperone polypeptides and polynucleotides are useful in the diagnosis,
CC   treatment and prevention of reproductive (e.g. prolactin production,
CC   infertility, endometrial or ovarian tumour, cancer of the breast,
CC   prostate or testis, Peyronie's disease), eye (e.g. conjunctivitis,
CC   keratitis, iritis, glaucoma), neuromuscular, metabolic (e.g. Zellweger
CC   syndrome, Addison's disease, cystic fibrosis), and autoimmune and
CC   inflammatory disorders (e.g. systemic lupus erythematosus, acquired
CC   immunodeficiency syndrome (AIDS), asthma, atherosclerosis), infectious
CC   or viral diseases, and cell proliferative disorders. Chaperone
CC   polynucleotides may be used for somatic or germline gene therapy, to
CC   detect and quantify gene expression in biopsied tissues in which
CC   expression is correlated with disease.
XX
SQ   Sequence 1147 BP; 287 A; 278 C; 314 G; 268 T; 0 other;

Query Match   100.0%;   Score 1147;   DB 22;   Length 1147;
Best Local Similarity 100.0%;   Pred. No 0;
Matches 1147;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   1 GCCTTGGGTCAAGCAGAGAATATTAATAGGCAGGGAATGCACCTGTAGCTACTGGGCGCTA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GCCTTGGGTCAAGCAGAGAATATTAATAGGCAGGGAATGCACCTGTAGCTACTGGGCGCTA 60

Qy   61 CTGCCAGCCTGAAGAGGAAGTGGCTCGACTCTTGACCATGCTGGGGTTCCTGAGGATGA 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CTGCCAGCCTGAAGAGGAAGTGGCTCGACTCTTGACCATGCTGGGGTTCCTGAGGATGA 120

Qy   121 GCTAAACCCTTTCCATGTACTGGGGGTTGAGGCCACAGCATCAGATGTTGAACTGAAGAA 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GCTAAACCCTTTCCATGTACTGGGGGTTGAGGCCACAGCATCAGATGTTGAACTGAAGAA 180

Qy   181 GGCCTATAGACAGCTGGCAGTGGCAGTGTGATGGTTCATCCTGACAAAAATCATCATCACCCGGCTGA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 GGCCTATAGACAGCTGGCAGTGGCAGTGTGATGGTTCATCCTGACAAAAATCATCATCACCCGGCTGA 240

Qy   241 GGAGGCCTTCAAGGTTTTGCGAGCAGCTTGGCAGCTTGGGACATTGTCAGCAATGTGCTGA... 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GGAGGCCTTCAAGGTTTTGCGAGCAGCTTGGCAGCTTGGGACATTGTCAGCAATGTGCTGAAAAGCGAAA 300

Qy   301 GGAGTATGAGATGAAACGAATGGCAGAGAATGAGCTGAGCCGGTCAGTAAATGAGTTTCT 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 GGAGTATGAGATGAAACGAATGGCAGAGAATGAGCTGAGCCGGTCAGTAAATGAGTTTCT 360

Qy   361 GTCCAAGCTGCAAGATGACCTCAAGGAGGAGCAATGAATACTATGAGTGTAGCCGATGCCA 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 GTCCAAGCTGCAAGATGACCTCAAGGAGGAGCAATGAATACTATGAGTGTAGCCGATGCCA 420

Qy   421 AGGAAAGCATAGGAGGTTTGAAAATGGACCGGGAACCTAAGAGTGCCAGATACTGTGCTGA 480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 AGGAAAGCATAGGAGGTTTGAAAATGGACCGGGAACCTAAGAGTGCCAGATACTGTGCTGA 480

Qy   481 GTGTAATAGGCTGCATCCTGCTGAGGAAGGAGACTTTGGGCAGAGTCAAGCATGTTGGG 540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 GTGTAATAGGCTGCATCCTGCTGAGGAAGGAGACTTTGGGCAGAGTCAAGCATGTTGGG 540

Qy   541 CCTCAAGATCACCTACTTTGCACTGATGATGAAAGGTGTATGACATCACAGAGTGGGC 600
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 CCTCAAGATCACCTACTTTGCACTGATGATGAAAGGTGTATGACATCACAGAGTGGGC 600

Qy   601 TGGATGCCAGCGTGGTGTAGTAGTATCCCCAGATACCCACAGAGTCCCTATCACATCTCATT 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 TGGATGCCAGCGTGGTGTAGTAGTATCCCCAGATACCCACAGAGTCCCTATCACATCTCATT 660

Qy   661 TGGTTCTCGGATTCCAGGCACCAGAGGGCGGGCAGAGAGAGCCACCCAGATGCCCTCCTGC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 TGGTTCTCGGATTCCAGGCACCAGAGGGCGGGCAGAGAGAGCCACCCAGATGCCCTCCTGC 720

Qy   721 TGATCTTCAGGATTTCTTGAGTCGGATCTTTCAAGTACCCCAGGGCAGATGCCCAATGG 780
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 TGATCTTCAGGATTTCTTGAGTCGGATCTTTCAAGTACCCCAGGGCAGATGCCCAATGG 780

Qy   781 GAACTTCTTTGCAGCCTCCTCAGCCTGCCCCGGAGCCTGCTGCAGCCTCTAAGCCACACAG 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 GAACTTCTTTGCAGCCTCCTCAGCCTGCCCCGGAGCCTGCTGCAGCCTCTAAGCCACAACAG 840

Qy   841 CACAGTACCCAAGGAGAAGCCAAACCTAAGCGCGGCCGAGAAAGTGAGGAGGCCCTTCCA 900
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 CACAGTACCCAAGGAGAAGCCAAACCTAAGCGCGGCCGAGAAAGTGAGGAGGCCCTTCCA 900

Qy   901 ACGTTGATGCCCCTTCTTCTTTCCTCAAATCAATGTCAGGGAGTCAAAAGGGCTGTAGCAC 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 ACGTTGATGCCCCTTCTTCTTTCCTCAAATCAATGTCAGGGAGTCAAAAGGGCTGTAGCAC 960

Qy   961 AGGATGGAGTTTGATTATCCCTCGTCTGTTGCCCAGTGGAGTGCAGTGGTGTGATCCTCAG 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961 AGGATGGAGTTTGATTATCCCTCGTCTGTTGCCCAGTGGAGTGCAGTGGTGTGATCCTCAG 1020

Qy   1021 TATTTTTTGAGATGGAGTCTTGCTCTGTCTCCCGGGTTCAAGCAATTCTCCCATCTCAGCTTCTCCCATCTCAGCGCCTCGAGTAG 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 TATTTTTTGAGATGGAGTCTTGCTCTGTCTCCCGGGTTCAAGCAATTCTCCCATCTCAGCTTCTCCCATCTCAGCGCCTCGAGTAG 1080

Qy   1081 CTTACTGCCAACCTCTGTCTCCCGGGTTCAAGCAATTCTCCCATCTCAGCTTCTCCCATCTCAGCGCCTCGAGTAG 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 CTTACTGCCAACCTCTGTCTCCCGGGTTCAAGCAATTCTCCCATCTCAGCTTCTCCCATCTCAGCGCCTCGAGTAG 1140

Qy   1141 CTGGGAT 1147
          |||||||
Db   1141 CTGGGAT 1147

RESULT 2
ABL55088
ID   ABL55088 standard; DNA; 2279 BP.
XX
AC   ABL55088;
XX
DT   08-OCT-2002 (first entry)
XX
DE   Human cDNA sequence #1 from clone HNTPBB2.
XX
KW   Human; HNTPB2; secreted protein; immunosuppressive; food preservative;
KW   antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;
KW   vasocropic; cerebroprotective; nootropic; neuroprotective; antibacterial;
KW   virucide; fungicide; ophthalmological; vulnerary; gene therapy; ELISA;
KW   radioimmunoassay; enzyme linked immunosorbent assay; autoimmune disease;
KW   hyperproliferative disorder; cardiovascular disorder; angiogenesis;
KW   cerebrovascular disorder; nervous system disorder; ocular disorder;
KW   wound healing; food additive; ss.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   CDS   36..1274
FT   /*tag= a
FT   sig_peptide   36..38
FT   /*tag= b
FT   mat_peptide   39..1271
FT   /*tag= c
XX

OM nucleic - nucleic search, using sw model

Copyright   GenCore version 5.1.6
          (c) 1993 - 2003 Compugen Ltd.

Run on: December 1, 2003, 07:30:31 ; Search time 379 Seconds
                    (without alignments)
                    8169.539 Million cell updates/sec

Title: US-10-049-742-22
Perfect score: 1147
Sequence: 1 gcctggggtcaagcagaata.........gcctcctgagtagctgggat 1147

Scoring table: IDENTITY.NUC
               Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   N_Geneseq 19Jun03;*
   1: /SIDS17/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
   2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
   3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
   4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
   5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
   6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
   7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
   8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
   9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
  10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
  11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
  12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
  13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
  14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
  15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
  16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
  17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
  18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
  19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
  20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
  21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
  22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
  23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
  24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
  25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1147	100.0	1147	22	AAF54994	Nucleotide sequenc
2	1145.4	99.9	2279	24	ABL55088	Human cDNA sequenc
3	1142.2	99.6	3286	24	ABL56700	Nucleotide sequenc
4	743	64.8	754	20	AAZ34492	Human LYST interac
5	743	64.8	754	24	ABX04184	Human mRNA differe
6	743	64.8	754	24	ABS76460	cDNA encoding huma
7	628.2	54.8	747	22	AAI96802	Human neuroblastom
8	453	39.5	7453	22	AAS30481	DNA encoding novel
9	453	39.5	7453	22	AAL06261	Human reproductive
10	272.8	23.8	884	24	ABL55114	Human cDNA sequenc
11	251.4	21.9	12119	22	AAC86936	Nucleotide sequenc
12	251.4	21.9	12578	21	AAZ36196	Nucleotide sequenc
13	251.4	21.9	12734	24	ABA95615	Chimeric BVDV/HCV
14	251.4	21.9	12842	21	AAZ36203	Nucleotide sequenc
15	251.4	21.9	13198	21	AAZ36211	Nucleotide sequenc
16	251.4	21.9	14078	21	AAA38807	Plasmid pBVDdN1, c
17	251.4	21.9	14578	21	AAZ48136	Bovine viral diarr
18	251.4	21.9	15065	21	AAZ36195	Nucleotide sequenc
19	251.4	21.9	16622	21	AAZ36212	Nucleotide sequenc
20	245.2	21.4	256	24	ABQ56856	Human colon cancer
21	239.4	20.9	11674	21	AAZ36210	Nucleotide sequenc
22	234.6	20.5	339	24	ABL85522	Human ovarian canc
23	145	12.6	2277	22	AAK68770	Human immune/haema
24	120	10.5	565	23	ABV60340	Human prostate exp
25	119	10.4	330	16	AAT24617	Human gene signatu
26	116.8	10.2	28001	25	ABS54410	Human hydroxymethy
27	116.8	10.2	28001	25	ABX93300	Gene encoding huma
28	116.4	10.1	3135	23	ABL25653	Drosophila melanog
29	115.6	10.1	1811	22	AAH14407	Human cDNA sequenc
30	115.6	10.1	3132	22	AAH18382	Human cDNA sequenc
31	113.6	9.9	498	24	ABK45936	cDNA encoding colo
32	113.2	9.9	4941	22	AAI98901	Human excretory re
33	113.2	9.9	4941	22	AAI63251	Human kidney relat
34	113.2	9.9	160820	25	ABQ76673	Androgen receptor
35	112.8	9.8	2147	23	ABK42851	Genomic sequence #
36	112.8	9.8	32174	22	ABA15665	Human nervous syst
37	112.8	9.8	32174	22	ABA19477	Human nervous syst
38	112.8	9.8	32174	22	ABA20359	Human nervous syst
39	112.8	9.8	32174	22	ABA21505	Human nervous syst
40	112.8	9.8	32174	22	AAL36280	Human musculoskele
41	112.8	9.8	32174	22	AAS32655	Human genomic DNA
42	112.8	9.8	32174	22	AAS34994	Human DNA for a no
43	112.8	9.8	32174	22	AAS30638	DNA encoding novel
44	112.8	9.8	32174	22	AAL03792	Human reproductive
45	112.8	9.8	32174	22	AAL07447	Human reproductive

ALIGNMENTS

RESULT 1
AAF54994
ID   AAF54994 standard; DNA; 1147 BP.
XX
AC   AAF54994;
XX
DT   15-MAY-2001 (first entry)
XX
DE   Nucleotide sequence of a human chaperone polypeptide.
XX
KW   Human; chaperone polypeptide; reproductive disease; prolactin production;
KW   infertility; tumour; cancer; Peyronie's disease; eye disorder; glaucoma;
KW   conjunctivitis; keratitis; neuromuscular disorder; cystic fibrosis;
KW   metabolic disorder; Zellweger syndrome; Addison's disease; iritis;
KW   autoimmune disorder; inflammatory disorder; systemic lupus erythematosus;
KW   acquired immunodeficiency syndrome; AIDS; asthma; atherosclerosis;
KW   cell proliferative disorder; gene therapy; ss.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   CDS   98..907
FT   /*tag= a
FT   /product= "chaperone protein"
XX
PN   WO200109178-A2.
XX
PD   08-FEB-2001.
XX
PF   03-AUG-2000; 2000WO-US21313.
XX

```
                    GenCore version 5.1.6
              Copyright (c) 1993 - 2003  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:   December  1, 2003, 07:30:31 ; Search time 90 Seconds
                                (without alignments)
                                5625.183 Million cell updates/sec

Title:     US-10-049-742-22
Perfect score:  1147
Sequence:  1 gccttggtcaagcagaata.........gcctcctgagtagctggat 1147

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:  569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters:   1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :  Issued_Patents_NA:*
            1: /cgn2_6/pcodata/2/ina/5A_COMB.seq:*
            2: /cgn2_6/pcodata/2/ina/5B_COMB.seq:*
            3: /cgn2_6/pcodata/2/ina/6A_COMB.seq:*
            4: /cgn2_6/pcodata/2/ina/6B_COMB.seq:*
            5: /cgn2_6/pcodata/2/ina/PCTUS_COMB.seq:*
            6: /cgn2_6/pcodata/2/ina/backfiles1.seq:*

         Pred. No. is the number of results predicted by chance to have a
         score greater than or equal to the score of the result being printed,
         and is derived by analysis of the total score distribution.

                              SUMMARIES
```

| | | % Query | | | | | |
| Result | Score | Match | Match | Length | DB | ID | | Description |
No.								
1	251.4	21.9	12734	4	US-09-344-456-1		Sequence 1, Appli
2	251.4	21.9	14078	3	US-09-433-262-1		Sequence 1, Appli
3	251.4	21.9	14078	3	US-09-702-330-1		Sequence 1, Appli
4	251.4	21.9	14578	3	US-09-859-694-1		Sequence 1, Appli
5	116.8	10.2	28001	4	US-09-819-993-3		Sequence 3, Appli
6	112.4	9.8	174493	4	US-09-804-471A-3		Sequence 3, Appli
c 7	110.4	9.6	43069	4	US-09-292-542A-1		Sequence 1, Appli
c 8	109.8	9.6	74962	4	US-09-685-853A-1		Sequence 3, Appli
9	107	9.3	955	4	US-09-641-638-4		Sequence 4, Appli
c 10	107	9.3	955	4	US-09-641-638-5		Sequence 5, Appli
c 11	107	9.3	955	4	US-09-641-638-6		Sequence 6, Appli
c 12	107	9.3	955	4	US-09-641-638-7		Sequence 7, Appli
c 13	107	9.3	1000	4	US-09-641-638-650		Sequence 650, App
14	106	9.2	2871	4	US-09-489-847-111		Sequence 111, App
15	106	9.2	2876	4	US-09-489-847-22		Sequence 22, Appli
c 16	106	9.2	43950	4	US-09-735-934A-3		Sequence 3, Appli
c 17	106	9.2	43950	4	US-10-060-332-3		Sequence 3, Appli
c 18	106	9.2	174493	4	US-09-804-471A-3		Sequence 3, Appli
19	105.2	9.2	31208	4	US-09-852-067-3		Sequence 3, Appli
c 20	104.8	9.1	2839	3	US-08-468-856B-5		Sequence 5, Appli
c 21	104.8	9.1	2839	3	US-08-468-859A-5		Sequence 5, Appli
22	104.6	9.1	562	3	US-09-385-982-442		Sequence 442, App
23	104.6	9.1	606	3	US-09-385-982-413		Sequence 413, App
24	104.6	9.1	632	3	US-09-385-982-177		Sequence 177, App
25	104.6	9.1	657	3	US-09-385-982-335		Sequence 335, App
26	104.4	9.1	653	3	US-09-385-982-324		Sequence 324, App
c 27	104.4	9.1	59065	4	US-09-813-817-3		Sequence 3, Appli
c 28	104.4	9.1	59065	4	US-09-978-197-3		Sequence 3, Appli
c 29	104.2	9.1	72604	4	US-09-228-992-7		Sequence 7, Appli
c 30	104.2	9.1	72604	4	US-09-657-474-7		Sequence 7, Appli
31	104	9.1	3507	1	US-08-832-883-67		Sequence 67, Appl
32	104	9.1	3507	2	US-08-832-877-67		Sequence 67, Appl
33	104	9.0	162450	4	US-09-345-882-1		Sequence 1, Appli
34	103.8	9.0	63588	4	US-09-873-404-3		Sequence 3, Appli
c 35	103	9.0	111282	4	US-09-754-250-3		Sequence 3, Appl
36	102.8	9.0	1002	4	US-09-641-638-578		Sequence 578, App
c 37	102.8	9.0	1184	4	US-09-489-847-76		Sequence 76, Appl
38	102.8	9.0	5789	3	US-09-242-948-3		Sequence 3, Appli
39	102.8	8.9	26664	4	US-09-564-805-28		Sequence 28, Appli
40	102.4	8.9	62804	4	US-09-800-960-3		Sequence 3, Appli
41	102.4	8.9	80246	3	US-09-078-294-4		Sequence 4, Appli
42	102.4	8.9	80595	3	US-09-078-294-3		Sequence 3, Appli
c 43	102.4	8.9	148567	4	US-09-801-876B-3		Sequence 3, Appli
c 44	102.2	8.9	3001	4	US-09-539-333D-187		Sequence 187, App
45	102.2	8.9	18073	3	US-09-078-294-12		Sequence 12, App

```
                              ALIGNMENTS

RESULT 1
US-09-344-456-1
; Sequence 1, Application US/09344456A
; Patent No. 6326137
; GENERAL INFORMATION:
;   APPLICANT: Hong, Zhi
;   APPLICANT: Lai, Vicky C.H.
;   APPLICANT: Lau, Jonhson Y.N.
;   TITLE OF INVENTION: HEPATITIS C VIRUS PROTEASE-DEPENDENT CHIMERIC
;   TITLE OF INVENTION: PESTIVIRUS
;   FILE REFERENCE: IN01038
;   CURRENT APPLICATION NUMBER: US/09/344,456A
;   CURRENT FILING DATE: 1999-06-25
;   NUMBER OF SEQ ID NOS: 4
;   SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 12734
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
; FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Chimeric
;   OTHER INFORMATION: Pestivirus
US-09-344-456-1

        Query Match             21.9%;  Score 251.4;  DB 4;  Length 12734;
        Best Local Similarity   94.2%;  Pred. No. 1.5e-68;
        Matches  261; Conservative    0; Mismatches   16; Indels    0;  Gaps    0;

Qy        404 ATGTGTAGCCGATGCCAAGGAAGCATAGGAGGTTGAAATGACCGGGAACCTAAGAGT 463
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5150 ATGTGCAGCCGATGCCAAGGAAGCATAGGAGGTTTGAAATGACCGGGAACCTAAGAGT 5209

Qy        464 GCCAGATACTGTGCTGAGTGTAATAGGCTGCTGCATCCTGCTGAGGAGGAGACTTTTGGGCA 523
              |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db       5210 GCCAGATACTGTGCTGAGTGTAATAGGCTGCTGCATCCTGCTGAGGAGGTGACTTTTGGGCA 5269

Qy        524 GAGTCAAGCACATGTTGGGCCTCAAGATCACCTACTTTGCACTGATGGATGGAAAGGTGTAT 583
              |||||||||||||||||||||||||||| |||||||||||| ||||||||||||||||||||
Db       5270 GAGTCGAGCACATGTTGGGCCTCAAAATCACCTACTTTTGCGCTGATGGATGGATGGAAAGGTGTAT 5329

Qy        584 GACATCACAGAGTGGGCTGGCTGGATGCCAGCGTGTAGGTATCTCCCAGATACCCACAGAGTC 643
              ||||| |||||||||||||||||||||||||||||| |||||||||| ||||||||||||||
Db       5330 GATATCACAGAGTGGGCTGGGCTGGATGCCAGCGTGTGGGGAATCTCCCAGATACCCACAGAGTC 5389

Qy        644 CCCTATCACATCTCATTTGGTTCTCCGGATTCCAGGCA 680
              |||||||||||||||||||||||| ||||||||| ||||
Db       5390 CCTTGTCACATCTCATTTGGTTCACGGATGCCTTTCA 5426

RESULT 2
```

US-09-433-262-1
; Sequence 1, Application US/09433262
; Patent No. 6168942
; GENERAL INFORMATION:
; APPLICANT: Cao, Xuemei
; APPLICANT: Sheppard, Mike
; TITLE OF INVENTION: ATTENUATED FORMS OF BOVINE VIRAL DIARRHEA VIRUS
; FILE REFERENCE: PC10435A
; CURRENT APPLICATION NUMBER: US/09/433,262
; CURRENT FILING DATE: 1999-11-04
; EARLIER APPLICATION NUMBER: 60/107,908
; EARLIER FILING DATE: 1998-11-10
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 14078
; TYPE: DNA
; ORGANISM: Bovine Viral Diarrhea Virus
US-09-433-262-1

Query Match   21.9%; Score 251.4; DB 3; Length 14078;
Best Local Similarity 94.2%; Pred. No. 1.6e-68;
Matches 261; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy   404 ATGTGTAGCCGATGCCAAGGAAAGCATAGGAGGTTTGAAATGGACCGGGAACCTAAGAGT 463
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 4532 ATGTGCAGCCGATGCCAAGGAAAGCATAGGAGGTTTGAAATGGACCGGGAACCTAAGAGT 4591

RESULT 3
US-09-702-330-1
; Sequence 1, Application US/09702330
; Patent No. 6410032
; GENERAL INFORMATION:
; APPLICANT: Cao, Xuemei
; APPLICANT: Sheppard, Mike
; TITLE OF INVENTION: ATTENUATED FORMS OF BOVINE VIRAL DIARRHEA VIRUS
; FILE REFERENCE: PC10435A
; CURRENT APPLICATION NUMBER: US/09/702,330
; CURRENT FILING DATE: 2000-10-31
; PRIOR APPLICATION NUMBER: 09/433,262
; PRIOR FILING DATE: 1999-11-04
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 14078
; TYPE: DNA
; ORGANISM: Bovine Viral Diarrhea Virus
US-09-702-330-1

Query Match   21.9%; Score 251.4; DB 4; Length 14078;
Best Local Similarity 94.2%; Pred. No. 1.6e-68;
Matches 261; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy   404 ATGTGTAGCCGATGCCAAGGAAAGCATAGGAGGTTTGAAATGGACCGGGAACCTAAGAGT 463
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 4532 ATGTGCAGCCGATGCCAAGGAAAGCATAGGAGGTTTGAAATGGACCGGGAACCTAAGAGT 4591

Qy   464 GCCAGATACTGTGCTGAGTGTAATAGGCTGCATCCTGCTGAGGAAGGAGACTTTGGGCA 523
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 4592 GCCAGATACTGTGCTGAGTGTAATAGGCTGCATCCTGCTGAGGAAGGAGACTTTGGGCA 4651

Qy   524 GAGTCAAGCATGTTGGGCCTCAAGATCACCTACTTTGCACTGATGGAAGGTGTAT 583
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 4652 GAGTCGAGCATGTTGGGCCTCAAAATCACCTACTTTGCGCTGATGGAAGGTGTAT 4711

Qy   584 GACATCACAGAGTGGGCTGGGCTGGATGCCAGCGTGGTAGTCTCCCCAGATACCCACAGAGTC 643
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 4712 GATATCACAGAGTGGGCTGGGCTGGATGCCAGCGTGGTAGTCTCCCCAGATACCCACAGAGTC 4771

Qy   644 CCCTATCACATCCATTGGTTCTCGGATTCCAGGCA 680
         ||||||||||||||||||||||||||||||||
Db 4772 CCTTGTCACATCCATTGGTTCACGGATGCCTTTCA 4808

RESULT 4
US-08-859-694-1
; Sequence 1, Application US/0859694A
; Patent No. 6001613
; GENERAL INFORMATION:
; APPLICANT: Donis, Ruben O.
; APPLICANT: Vassilev, Ventzislav B.
; TITLE OF INVENTION: A plasmid bearing a cDNA copy of the genome of bovine
; TITLE OF INVENTION: viral diarrhea virus, chimeric derivations thereof, and
; TITLE OF INVENTION: method of producing an infectious bovine viral diarrhea
; TITLE OF INVENTION: virus using said plasmid
; FILE REFERENCE: UNVN51110
; CURRENT APPLICATION NUMBER: US/08/859,694A
; CURRENT FILING DATE: 1997-05-21
; EARLIER APPLICATION NUMBER: 60/018,246
; EARLIER FILING DATE: 1996-05-24
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 14578
; TYPE: DNA
; ORGANISM: bovine viral diarrhea virus
US-08-859-694-1

Query Match   21.9%; Score 251.4; DB 3; Length 14578;
Best Local Similarity 94.2%; Pred. No. 1.6e-68;
Matches 261; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy   404 ATGTGTAGCCGATGCCAAGGAAAGCATAGGAGGTTTGAAATGGACCGGGAACCTAAGAGT 463
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 5032 ATGTGCAGCCGATGCCAAGGAAAGCATAGGAGGTTTGAAATGGACCGGGAACCTAAGAGT 5091

Qy   464 GCCAGATACTGTGCTGAGTGTAATAGGCTGCATCCTGCTGAGGAAGGAGACTTTGGGCA 523
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 5092 GCCAGATACTGTGCTGAGTGTAATAGGCTGCATCCTGCTGAGGAAGGAGTGACTTTGGGCA 5151

Qy   524 GAGTCAAGCATGTTGGGCCTCAAGATCACCTACTTTGCACTGATGGAAGGTGTAT 583
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 5152 GAGTCGAGCATGTTGGGCCTCAAAATCACCTACTTTGCGCTGATGGAAGGTGTAT 5211

Qy   584 GACATCACAGAGTGGGCTGGGCTGGATGCCAGCGTGGTAGTCTCCCCAGATACCCACAGAGTC 643
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 5212 GATATCACAGAGTGGGCTGGGCTGGATGCCAGCGTGGTAGTCTCCCCAGATACCCACAGAGTC 5271

Qy   644 CCCTATCACATCCATTGGTTCTCGGATTCCAGGCA 680
         ||||||||||||||||||||||||||||||||
Db 5272 CCTTGTCACATCCATTGGTTCACGGATGCCTTTCA 5308

RESULT 5
US-09-819-993-3/c
; Sequence 3, Application US/09819993
; Patent No. 6436692
; GENERAL INFORMATION:
; APPLICANT: GONG, Fangcheng et al.

```
;  TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
;  TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
;  TITLE OF INVENTION: THEREOF
;  FILE REFERENCE: CL001195
;  CURRENT APPLICATION NUMBER: US/09/819,993
;  CURRENT FILING DATE: 2001-03-29
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: FastSEQ for Windows Version 4.0
;  SEQ ID NO 3
;  LENGTH: 28001
;  TYPE: DNA
;  ORGANISM: Human
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (1)..(28001)
;  OTHER INFORMATION: n = A,T,C or G
US-09-819-993-3

Qy          1004 CTGAATCTTTTCTTTTTTATTTTGAGATGGAGTCTTGCTCTGTTGCCCAGCTGGAGTG 1063
                      | ||||||| |||||||||  |||||||||||||||||||||||||||||||||
Db         23843 CTTTTTTTTTTTTTTTTTTTTTTTGAGATGGAATCTTGCTCTCTGTCGCCAGGCTGGAGTG 23784

Qy          1064 CAGTGGTGTGATCTCAGCTTACTGCAACCTCTGTCTCCCGGGTTCAAGCAATTCTCCCAT 1123
                 |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db         23783 CAGTGGTGTGATCTCAGCTTACTGCAACCTCTGACTCCCTGTTCAAGCCATTCTCCTGA 23724

Qy          1124 CTCAGCCCCTGAGTAGCTGGGAT 1147
                 |||||||||||||||||| ||
Db         23723 CTCAGCCTCCTGAGTAGCTGGAAT 23700

RESULT 6
US-09-804-471A-3
;  Sequence 3, Application US/09804471A
;  Patent No. 6479269
;  GENERAL INFORMATION:
;  APPLICANT: WEBSTER, Marion et al
;  TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
;  TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
;  TITLE OF INVENTION: THEREOF
;  FILE REFERENCE: CL001164
;  CURRENT APPLICATION NUMBER: US/09/804,471A
;  CURRENT FILING DATE: 2001-03-13
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
;  SEQ ID NO 3
;  LENGTH: 174493
;  TYPE: DNA
;  ORGANISM: Human
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (1)..(174493)
;  OTHER INFORMATION: n = A,T,C or G
US-09-804-471A-3

Qy          1009 TCTTTTTCTTTTTTATTTTTTGAGATGGAGTCTTGCTCTGTTGCCCAGCTGGAGTGCAGTG 1068
                      ||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db        119174 TTTTTTTTTTTTTTTTTTTTTTAGAAGGAGTCTCGCTCTGTTGCCAGGCTGGAGTGCAGTG 119233

Qy          1069 GTGTGATCTCAGCTTACTGCAACCTCTGTCTCCCGGGTTCAAGCAATTCTCCCATCTTCAG 1128
                 |||||||||||||||||||||||||||||  |||||||||||||||||||||||||||||
Db        119234 GTGTGATCTCAGCTTACTGCAACCTCTGCCTCCCGGGTTCAAGTGATTCTCCTGCCTCAG 119293

Qy          1129 CCTCCTGAGTAGCTGGGA 1146
                 ||||||||||||||
```

Db   119294 CCTCCCGAGTAGCTGGGA 119311

```
RESULT 7
US-09-292-542A-1/c
;  Sequence 1, Application US/09292542A
;  Patent No. 6531279
;  GENERAL INFORMATION:
;  APPLICANT: Blumenfeld, Marta
;  APPLICANT: Chumakov, Ilya
;  APPLICANT: Bougueleret, Lydie
;  TITLE OF INVENTION: Genomic Sequence Of The 5-Lipoxygenase-Activating Protein (FLAP)
;  TITLE OF INVENTION: Polymorphic Markers Thereof And Methods For Detection Of Asthma.
;  FILE REFERENCE: GENSET.026A
;  CURRENT FILING DATE: 1999-04-15
;  CURRENT APPLICATION NUMBER: US/09/292,542A
;  PRIOR APPLICATION NUMBER: US 60/081893
;  PRIOR FILING DATE: 1998-04-15
;  PRIOR APPLICATION NUMBER: US 60/091314
;  PRIOR FILING DATE: 1998-06-30
;  PRIOR APPLICATION NUMBER: US 60/123406
;  PRIOR FILING DATE: 1999-03-08
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: Patent.pm
;  SEQ ID NO 1
;  LENGTH: 43069
;  TYPE: DNA
;  ORGANISM: homo sapiens
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: 1..7708
;  OTHER INFORMATION: potential 5'regulatory region
;  NAME/KEY: misc_feature
;  LOCATION: 36604..43069
;  OTHER INFORMATION: potential 3'regulatory region
;  NAME/KEY: exon
;  LOCATION: 7709..7852
;  NAME/KEY: exon
;  OTHER INFORMATION: exon1
;  NAME/KEY: exon
;  LOCATION: 16236..16335
;  OTHER INFORMATION: exon2
;  NAME/KEY: exon
;  LOCATION: 24227..24297
;  OTHER INFORMATION: exon3
;  NAME/KEY: exon
;  LOCATION: 28133..28214
;  OTHER INFORMATION: exon4
;  NAME/KEY: exon
;  LOCATION: 36128..36605
;  OTHER INFORMATION: exon5
;  NAME/KEY: misc_feature
;  LOCATION: 7783..7785
;  OTHER INFORMATION: ATG
;  NAME/KEY: misc_feature
;  LOCATION: 36288..36290
;  OTHER INFORMATION: stop : TAA
;  NAME/KEY: polyA_signal
;  LOCATION: 36581..36586
;  OTHER INFORMATION: AATAAA
;  NAME/KEY: misc_feature
;  LOCATION: 7008..8116
;  OTHER INFORMATION: homology with sequence in ref genbank : M60470
;  NAME/KEY: misc_feature
;  LOCATION: 15995..16549
;  OTHER INFORMATION: homology with sequence in ref genbank : M63259
;  NAME/KEY: misc_feature
;  LOCATION: 24059..24597
;  OTHER INFORMATION: homology with sequence in ref genbank : M63260
;  NAME/KEY: misc_feature
;  LOCATION: 27873..28412
;  OTHER INFORMATION: homology with sequence in ref genbank : M63261
;  NAME/KEY: misc_feature
```

Query Match   10.2%;   Score 116.8; DB 4; Length 28001;
Best Local Similarity 88.2%;   Pred. No. 6.9e-26;
Matches 127; Conservative   0; Mismatches 17; Indels   0; Gaps   0;

Query Match   9.8%;   Score 112.4; DB 4; Length 174493;
Best Local Similarity 88.4%;   Pred. No. 4.9e-24;
Matches 122; Conservative   0; Mismatches 16; Indels   0; Gaps   0;

Db   119294 CCTCCCGAGTAGCTGGGA 119311

LOCATION: 35977..36926
OTHER INFORMATION: homology with sequence in ref genbank : M63262
NAME/KEY: misc_feature
LOCATION: 7613
OTHER INFORMATION: diverging nucleotide deletion of a A in ref : M60470
NAME/KEY: misc_feature
LOCATION: 16347
OTHER INFORMATION: diverging nucleotide G in ref : M63259
NAME/KEY: misc_feature
LOCATION: 16348
OTHER INFORMATION: diverging nucleotide A in ref : M63259
NAME/KEY: misc_feature
LOCATION: 24060
OTHER INFORMATION: diverging nucleotide deletion of a G in ref : M63260
NAME/KEY: misc_feature
LOCATION: 24067
OTHER INFORMATION: diverging nucleotide deletion of a G in ref : M63260
NAME/KEY: misc_feature
LOCATION: 27903
OTHER INFORMATION: diverging nucleotide deletion of a C in ref : M63261
NAME/KEY: misc_feature
LOCATION: 28327
OTHER INFORMATION: diverging nucleotide deletion of a G in ref : M63261
NAME/KEY: misc_feature
LOCATION: 3851..4189
NAME/KEY: misc_feature
LOCATION: 10-517
NAME/KEY: misc_feature
LOCATION: 4120..4390
NAME/KEY: misc_feature
LOCATION: 10-518
NAME/KEY: misc_feature
LOCATION: 4373..4792
NAME/KEY: misc_feature
LOCATION: 10-253
NAME/KEY: misc_feature
LOCATION: 4814..5043
NAME/KEY: misc_feature
LOCATION: 10-499
NAME/KEY: misc_feature
LOCATION: 4956..5422
NAME/KEY: misc_feature
LOCATION: 10-500
NAME/KEY: misc_feature
LOCATION: 5524..5996
NAME/KEY: misc_feature
LOCATION: 10-522
NAME/KEY: misc_feature
LOCATION: 6218..6672
NAME/KEY: misc_feature
LOCATION: 10-503
NAME/KEY: misc_feature
LOCATION: 6522..6790
NAME/KEY: misc_feature
LOCATION: 10-504
NAME/KEY: misc_feature
LOCATION: 7120..7574
NAME/KEY: misc_feature
LOCATION: 10-204
NAME/KEY: misc_feature
LOCATION: 7513..7933
NAME/KEY: misc_feature
LOCATION: 10-32
NAME/KEY: misc_feature
LOCATION: 36020..36465
NAME/KEY: misc_feature
LOCATION: 10-36
NAME/KEY: misc_feature
LOCATION: 36318..36669
NAME/KEY: misc_feature
LOCATION: 10-498
NAME/KEY: misc_feature
LOCATION: 16014..16533
NAME/KEY: misc_feature
LOCATION: 10-33
NAME/KEY: misc_feature
LOCATION: 38441..38840
NAME/KEY: misc_feature
LOCATION: 12-629
NAME/KEY: misc_feature
LOCATION: 24072..24425
NAME/KEY: misc_feature
LOCATION: 10-34
NAME/KEY: misc_feature
LOCATION: 27978..28401
NAME/KEY: misc_feature
LOCATION: 10-35
NAME/KEY: misc_feature
LOCATION: 42223..42749
NAME/KEY: misc_feature

OTHER INFORMATION: 12-628 complement
NAME/KEY: allele
LOCATION: 3950
OTHER INFORMATION: 10-517-100 : polymorphic base S
NAME/KEY: allele
LOCATION: 4243
OTHER INFORMATION: 10-518-125 : polymorphic base K
NAME/KEY: allele
LOCATION: 4312
OTHER INFORMATION: 10-518-194 : polymorphic base R
NAME/KEY: allele
LOCATION: 4490
OTHER INFORMATION: 10-253-118 : polymorphic base R
NAME/KEY: allele
LOCATION: 4670
OTHER INFORMATION: 10-253-298 : polymorphic base S
NAME/KEY: allele
LOCATION: 4687
OTHER INFORMATION: 10-253-315 : polymorphic base Y
NAME/KEY: allele
LOCATION: 4968
OTHER INFORMATION: 10-499-155 : polymorphic base R
NAME/KEY: allele
LOCATION: 5140
OTHER INFORMATION: 10-500-185 : polymorphic base Y
NAME/KEY: allele
LOCATION: 5213
OTHER INFORMATION: 10-500-258 : polymorphic base K
NAME/KEY: allele
LOCATION: 5364
OTHER INFORMATION: 10-500-410 : polymorphic base R
NAME/KEY: allele
LOCATION: 5594
OTHER INFORMATION: 10-522-71 : polymorphic base R
NAME/KEY: allele
LOCATION: 6370
OTHER INFORMATION: 10-503-159 : polymorphic base K
NAME/KEY: allele
LOCATION: 6693
OTHER INFORMATION: 10-504-172 : polymorphic base W
NAME/KEY: allele
LOCATION: 6763
OTHER INFORMATION: 10-504-243 : polymorphic base M
NAME/KEY: allele
LOCATION: 7445
OTHER INFORMATION: 10-204-326 : polymorphic base R
NAME/KEY: allele
LOCATION: 7870
OTHER INFORMATION: 10-32-357 : polymorphic base M
NAME/KEY: allele
LOCATION: 16288
OTHER INFORMATION: 10-33-175 : polymorphic base Y
NAME/KEY: allele
LOCATION: 16347
OTHER INFORMATION: 10-33-234 : polymorphic base M
NAME/KEY: allele
LOCATION: 16383
OTHER INFORMATION: 10-33-270 : polymorphic base R
NAME/KEY: allele

```
Query Match          9.6%;  Score 110.4;  DB 4;  Length 43069;
Best Local Similarity   85.4%;  Pred. No. 9.2e-24;
Matches 123; Conservative   0; Mismatches  21; Indels   0; Gaps   0;

Qy    1004 CTGAATCTTTTCTTTTATTTTTGAGATGGAGTCTTGCTCTGTTGCCCAGCTGGAGTG 1063
           ||||||  |||||| ||||||||||||| ||||||||| ||||||||||||||||||
Db   34796 CTTTTTCTTTTCTTTTTTTTTTTTGAGACTGAGTCAAGCTCTATTGCCCAGCTGGAGTG 34737

Qy    1064 CAGTGGTGTGATCTCAGCTTACTGCAACCTCCGTCTCCCGGGTTCAAGCAATTCTCCCAT 1123
           |||||  |||||||||||| |||||||||||||| |||||||||||||| ||||||
Db   34736 CAGTGCACGATCTCAGATCACTGCAACCTCCGCCTCCCGGGTTCAAGCGATTCTCCTGC 34677

Qy    1124 CTCAGCCTCCTGAGTAGCTGGGAT 1147
```

```
Db    34676 CTCAGCCTCCTGAGTAGCTGGGAT 34653
            ||||||||||||||||||||||||
```

RESULT 8
US-09-685-853A-3
; Sequence 3, Application US/09685853A
; Patent No. 6479270
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui et al.
; TITLE OF INVENTION: ISOLATED HUMAN PHOSPHATASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PHOSPHATASE PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000871
; CURRENT APPLICATION NUMBER: US/09/685,853A
; CURRENT FILING DATE: 2002-05-06
; PRIOR APPLICATION NUMBER: 60/182,194
; PRIOR FILING DATE: 2000-02-14
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 74962
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(74962)
; OTHER INFORMATION: n = A,T,C or G
US-09-685-853A-3

Query Match   9.6%;   Score 109.8; DB 4; Length 74962;
Best Local Similarity   87.6%;   Pred. No. 2e-23;
Matches 120; Conservative   0; Mismatches 17; Indels 0; Gaps 0;

```
Qy  1011 TTTTTCTTTTATTTTTGAGATGGAGTCTTGCTCTGTTGCCCAGCTGGAGTGCAGTGGT 1070
         ||||| || ||||||||||| |||||||| || ||||| ||||||||||||||||||||
Db 65291 TTTTTTTGTTTTGTTTTGAGATGGAGTCTTGCTCTGTCGCCAGGCTGGAGTGCAGTGGT 65350

Qy  1071 GTGATCTCAGCTTACTGCAACCTCTGTCTCCCGGGTTCAAGCAATTCTCCCATCTCAGCC 1130
         ||||||||| ||| ||| |||||||| | |||||||||||||| || ||| ||||||||
Db 65351 GTGATCTCAGCTCACTGCAACCTTTGCCTCCCGGGTTCAAGCGATTCTCGCCTCAGCT 65410

Qy  1131 TCCTGAGTAGCTGGGAT 1147
         |||||||||||||||||
Db 65411 TCCTGAGTAGCTGGGAT 65427
```

RESULT 9
US-09-641-638-4/c
; Sequence 4, Application US/09641638
; Patent No. 6432648
; GENERAL INFORMATION:
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Chumakov, Ilya
; APPLICANT: Cohen, Annick
; TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
; TITLE OF INVENTION: GENES INVOLVED IN ARACHIDONIC ACID METABOLISM
; FILE REFERENCE: GENSET.051CP1
; CURRENT APPLICATION NUMBER: US/09/641,638
; CURRENT FILING DATE: 2000-08-16
; PRIOR APPLICATION NUMBER: US 09/502,330
; PRIOR FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: US 60/133,200
; PRIOR FILING DATE: 1999-05-07
; PRIOR APPLICATION NUMBER: US 09/275,267
; PRIOR FILING DATE: 1999-03-23
; PRIOR APPLICATION NUMBER: US 60/119,917
; PRIOR FILING DATE: 1999-02-12
; NUMBER OF SEQ ID NOS: 1304
; SOFTWARE: Patent.pm
; SEQ ID NO 4
; LENGTH: 955

; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: allele
; LOCATION: 478
; OTHER INFORMATION: 10-499-155 : polymorphic base A or G
; NAME/KEY: misc_binding
; LOCATION: 458..477
; OTHER INFORMATION: 10-499-155.mis1, potential
; NAME/KEY: misc_binding
; LOCATION: 479..498
; OTHER INFORMATION: 10-499-155.mis2, potential complement
; NAME/KEY: primer_bind
; LOCATION: 324..343
; OTHER INFORMATION: upstream amplification primer
; NAME/KEY: primer_bind
; LOCATION: 536..553
; OTHER INFORMATION: downstream amplification primer, complement
; NAME/KEY: misc_binding
; LOCATION: 466..490
; OTHER INFORMATION: 10-499-155 potential probe
US-09-641-638-4

Query Match   9.3%;   Score 107; DB 4; Length 955;
Best Local Similarity   77.6%;   Pred. No. 1.2e-23;
Matches 142; Conservative   0; Mismatches 40; Indels 1; Gaps 1;

```
Qy   965 TGGAGTTGATTTATCCTCCTCCCCAACACCTAGGAACTGAATCTTTTCTTTTATT 1024
         |||||  || |||||||||||| || |||||||||||||||  ||| ||||||||||
Db   910 TGGAGCTAGAGAGCTTTCTCTCTCCCGCCCTTCCTCCTCTCCTCTCTTTTTTTTTTTTT 851

Qy  1025 TTTTGAGATGGAGTCTCGTCTCCCGGGTGCCCA-GCTGGAGTGCAGTGGTGTGATCTCAGCTT 1083
         |||||| ||||||||| |||||||||||| ||| |||||||||||||||||| ||||||||||
Db   850 TTTTGAGACGGAGTCTCGCTCTGTCGCCAGGCTGGAGTGCAGTGGCGGGATCTCAGCTC 791

Qy  1084 ACTGCAACCTCTGTCTCCCGGGTTCAAGCAATTCCCATCTCAGCCTCCTGAGTAGCTG 1143
         |||||| ||||| |||||||||||||| || |||| || ||||||||||||||||||||
Db   790 ACTGCAAGCTCTGCCTCCCGGGTTCACGCCATTCTCCGTCTCCTCAGCCTCCTCAGTAGCTG 731

Qy  1144 GGA 1146
         |||
Db   730 GGA 728
```

RESULT 10
US-09-641-638-5/c
; Sequence 5, Application US/09641638
; Patent No. 6432648
; GENERAL INFORMATION:
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Chumakov, Ilya
; APPLICANT: Cohen, Annick
; TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
; TITLE OF INVENTION: GENES INVOLVED IN ARACHIDONIC ACID METABOLISM
; FILE REFERENCE: GENSET.051CP1
; CURRENT APPLICATION NUMBER: US/09/641,638
; CURRENT FILING DATE: 2000-08-16
; PRIOR APPLICATION NUMBER: US 09/502,330
; PRIOR FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: US 60/133,200
; PRIOR FILING DATE: 1999-05-07
; PRIOR APPLICATION NUMBER: US 09/275,267
; PRIOR FILING DATE: 1999-03-23
; PRIOR APPLICATION NUMBER: US 60/119,917
; PRIOR FILING DATE: 1999-02-12
; NUMBER OF SEQ ID NOS: 1304
; SOFTWARE: Patent.pm
; SEQ ID NO 5
; LENGTH: 955
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:

NAME/KEY: allele
LOCATION: 478
OTHER INFORMATION: 10-520-256 : polymorphic base C or T
NAME/KEY: misc_binding
LOCATION: 458..477
OTHER INFORMATION: 10-520-256.mis1, potential
NAME/KEY: misc_binding
LOCATION: 479..498
OTHER INFORMATION: 10-520-256.mis2, potential complement
NAME/KEY: primer_bind
LOCATION: 479..498
OTHER INFORMATION: 10-520-256.mis2, potential complement
NAME/KEY: primer_bind
LOCATION: 294..310
OTHER INFORMATION: upstream amplification primer
NAME/KEY: primer_bind
LOCATION: 743..760
OTHER INFORMATION: downstream amplification primer, complement
NAME/KEY: misc_binding
LOCATION: 466..490
OTHER INFORMATION: 10-520-256 potential probe
US-09-641-638-5

Query Match   9.3%;   Score 107; DB 4;   Length 955;
Best Local Similarity 77.6%;   Pred. No. 1.2e-23;
Matches 142; Conservative   0; Mismatches 40; Indels   1; Gaps   1;

Qy   965 TGGAGTTTGATTTATCCCTCCTCCCCAACACCTAGGAACTGAATCTTTTTCTTTTTATT 1024
            ||||||| || ||||||||||||||| ||
Db   738 TGGAGCTAGAGAGCTTTCTCTCTCTCTCCCGCCCTTCCTCTCTTTTTTTTTTTTT 679

Qy 1025 TTTTGAGATGGAGTCTTGCTCTGTTGCCCA-GCTGGAGTGCAGTGGTGATCTCAGCTT 1083
            ||||||||||| ||||| |||||||||| |||||||||||||||||||||||||
Db   678 TTTTGAGACGGAGTCTTGCTCTGTTGCCCAGGCAGTCAGTGGCGATCTCAGCTC 619

Qy 1084 ACTGCAACCTCTGTCTCCCGGGTCAAGCAATTCTCCCATCTCAGCCTCTGAGTAGCTG 1143
            ||||||||||||| ||||||||||||||||| || |||||||||||||||||||
Db   618 ACTGCAAGCTCTGCCTCCCGGGTTCACGCCATTCCTGTCTCAGCCTCCTCAGTAGCTG 559

Qy 1144 GGA 1146
            |||
Db   558 GGA 556

RESULT 11
US-09-641-638-6/c
; Sequence 6, Application US/09641638
; Patent No. 6432648
; GENERAL INFORMATION:
;   APPLICANT: Blumenfeld, Marta
;   APPLICANT: Bougueleret, Lydie
;   APPLICANT: Chumakov, Ilya
;   APPLICANT: Cohen, Annick
;   TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
;   TITLE OF INVENTION: GENES INVOLVED IN ARACHIDONIC ACID METABOLISM
;   FILE REFERENCE: GENSET.051CP1
;   CURRENT APPLICATION NUMBER: US/09/641,638
;   CURRENT FILING DATE: 2000-08-16
;   PRIOR APPLICATION NUMBER: US 09/502,330
;   PRIOR FILING DATE: 2000-02-11
;   PRIOR APPLICATION NUMBER: US 60/133,200
;   PRIOR FILING DATE: 1999-05-07
;   PRIOR APPLICATION NUMBER: US 09/275,267
;   PRIOR FILING DATE: 1999-03-23
;   PRIOR APPLICATION NUMBER: US 60/119,917
;   PRIOR FILING DATE: 1999-02-12
;   NUMBER OF SEQ ID NOS: 1304
;   SOFTWARE: Patent.pm
; SEQ ID NO 6
;   LENGTH: 955
;   TYPE: DNA
;   ORGANISM: Homo Sapiens
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 478
;   OTHER INFORMATION: 10-500-258 : polymorphic base G or T

NAME/KEY: misc_binding
LOCATION: 458..477
OTHER INFORMATION: 10-500-258.mis1, potential
NAME/KEY: misc_binding
LOCATION: 479..498
OTHER INFORMATION: 10-500-258.mis2, potential complement
NAME/KEY: primer_bind
LOCATION: 221..237
OTHER INFORMATION: upstream amplification primer
NAME/KEY: primer_bind
LOCATION: 670..687
OTHER INFORMATION: downstream amplification primer, complement
NAME/KEY: misc_binding
LOCATION: 466..490
OTHER INFORMATION: 10-500-258 potential probe
US-09-641-638-6

Query Match   9.3%;   Score 107; DB 4;   Length 955;
Best Local Similarity 77.6%;   Pred. No. 1.2e-23;
Matches 142; Conservative   0; Mismatches 40; Indels   1; Gaps   1;

Qy   965 TGGAGTTTGATTTATCCCTCCTCCCCAACACCTAGGAACTGAATCTTTTTCTTTTTATT 1024
            ||||||| || ||||||||||||||| ||
Db   665 TGGAGCTAGAGAGCTTTCTCTCTCTCTCCCGCCCTTCCTCTCTTTTTTTTTTTTTT 606

Qy 1025 TTTTGAGATGGAGTCTTGCTCTGTTGCCCA-GCTGGAGTGCAGTGGTGATCTCAGCTT 1083
            ||||||||||| ||||| |||||||||| |||||||||||||||||||||||||
Db   605 TTTTGAGACGGAGTCTTGCTCTGTTGCCCAGGCAGTCAGTGGCGCGGATCTCAGCTC 546

Qy 1084 ACTGCAACCTCTGTCTCCCGGGTCAAGCAATTCTCCCATCTCAGCCTCTGAGTAGCTG 1143
            ||||||||||||| |||||||||||||||| || |||||||||||||||||||
Db   545 ACTGCAAGCTCTGCCTCCCGGGTTCACGCCATTCCTGTCTCAGCCTCCTCAGTAGCTG 486

Qy 1144 GGA 1146
            |||
Db   485 GGA 483

RESULT 12
US-09-641-638-7/c
; Sequence 7, Application US/09641638
; Patent No. 6432648
; GENERAL INFORMATION:
;   APPLICANT: Blumenfeld, Marta
;   APPLICANT: Bougueleret, Lydie
;   APPLICANT: Chumakov, Ilya
;   APPLICANT: Cohen, Annick
;   TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
;   TITLE OF INVENTION: GENES INVOLVED IN ARACHIDONIC ACID METABOLISM
;   FILE REFERENCE: GENSET.051CP1
;   CURRENT APPLICATION NUMBER: US/09/641,638
;   CURRENT FILING DATE: 2000-08-16
;   PRIOR APPLICATION NUMBER: US 09/502,330
;   PRIOR FILING DATE: 2000-02-11
;   PRIOR APPLICATION NUMBER: US 60/133,200
;   PRIOR FILING DATE: 1999-05-07
;   PRIOR APPLICATION NUMBER: US 09/275,267
;   PRIOR FILING DATE: 1999-03-23
;   PRIOR APPLICATION NUMBER: US 60/119,917
;   PRIOR FILING DATE: 1999-02-12
;   NUMBER OF SEQ ID NOS: 1304
;   SOFTWARE: Patent.pm
; SEQ ID NO 7
;   LENGTH: 955
;   TYPE: DNA
;   ORGANISM: Homo Sapiens
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 478
;   OTHER INFORMATION: 10-500-410 : polymorphic base A or G
;   NAME/KEY: misc_binding
;   LOCATION: 458..477
;   OTHER INFORMATION: 10-500-410.mis1, potential

NAME/KEY: primer_bind
LOCATION: 431..449
OTHER INFORMATION: upstream amplification primer
NAME/KEY: primer_bind
LOCATION: 885..903
OTHER INFORMATION: downstream amplification primer, complement
NAME/KEY: misc_binding
LOCATION: 489..513
OTHER INFORMATION: 10-522-71 potential probe
US-09-641-638-650:

```
Query Match            9.3%;    Score 107; DB 4;  Length 1000;
Best Local Similarity 77.6%;    Pred. No. 1.2e-23;
Matches  142;   Conservative    0;   Mismatches  40;   Indels   1;   Gaps   1;

Qy   965 TGGAGTTGATTATCCTCCTCCCCCAACACCTAGGAACTGAATCTTTTCTTTTATT 1024
             ||||| ||                         ||||||| || ||
Db   307 TGGAGCTAGAGAGCTTTCTCTCTCTCTCTTTTTTTTTTTTT  248

Qy  1025 TTTTGAGATGGAGTCTTGCTCTGTTGCCCA-GCTGGAGTGCAGTGGTGTGATCTCAGCTT 1083
             ||||||| |||||||||||||||||||| |||||||||||||||| | ||||||||| 
Db   247 TTTTGAGACGGAGTCTTGCTCTGTTGCCCAGGCTGGAGTGCAGTGGCGGGATCTCAGCTC  188

Qy  1084 ACTGCAACCTCTGTCTCCCGGGTTCAAGCAATTCTCCATCTCAGCCTCCTGAGTAGCTG 1143
             ||||||| ||||| ||||||||||| | ||||||| ||| |||||||||| |||||||
Db   187 ACTGCAAGCTCTGCCTCCCGGGTTCACGCCATTCTCCTGTCTCAGCCTCCTCAGTAGCTG  128

Qy  1144 GGA 1146
             |||
Db   127 GGA  125
```

RESULT 14
US-09-489-847-111/c
; Sequence 111, Application US/09489847
; Patent No. 6476195
; GENERAL INFORMATION:
; APPLICANT: Rosen et al
; TITLE OF INVENTION: 98 Human Secreted Proteins
; FILE REFERENCE: PZ031P1
; CURRENT APPLICATION NUMBER: US/09/489,847
; CURRENT FILING DATE: 2000-01-24
; EARLIER APPLICATION NUMBER: PCT/US99/17130
; EARLIER FILING DATE: 1999-07-29
; EARLIER APPLICATION NUMBER: 60/094,657
; EARLIER FILING DATE: 1998-07-30
; EARLIER APPLICATION NUMBER: 60/095,486
; EARLIER FILING DATE: 1998-08-05
; EARLIER APPLICATION NUMBER: 60/096,319
; EARLIER FILING DATE: 1998-08-12
; EARLIER APPLICATION NUMBER: 60/095,454
; EARLIER FILING DATE: 1998-08-06
; EARLIER APPLICATION NUMBER: 60/095,455
; EARLIER FILING DATE: 1998-08-06
; NUMBER OF SEQ ID NOS: 376
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 111
; LENGTH: 2871
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1234)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1259)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (128)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:

NAME/KEY: misc_binding
LOCATION: 479..498
OTHER INFORMATION: 10-500-410.mis2, potential complement
NAME/KEY: primer_bind
LOCATION: 70..86
OTHER INFORMATION: upstream amplification primer
NAME/KEY: primer_bind
LOCATION: 519..536
OTHER INFORMATION: downstream amplification primer, complement
NAME/KEY: misc_binding
LOCATION: 466..490
OTHER INFORMATION: 10-500-410 potential probe
US-09-641-638-7:

```
Query Match            9.3%;    Score 107; DB 4;  Length 955;
Best Local Similarity 77.6%;    Pred. No. 1.2e-23;
Matches  142;   Conservative    0;   Mismatches  40;   Indels   1;   Gaps   1;

Qy   965 TGGAGTTGATTATCCTCCTCCCCCAACACCTAGGAACTGAATCTTTTCTTTTATT 1024
             ||||| ||                       ||||||| || ||
Db   514 TGGAGCTAGAGAGCTTTCTCTCTCCCGGCYCTTCCTCTCTTTTTTTTTTTTT  455

Qy  1025 TTTTGAGATGGAGTCTTGCTCTGTTGCCCA-GCTGGAGTGCAGTGGTGTGATCTCAGCTT 1083
             ||||||| |||||||||||||||||||| |||||||||||||||| | ||||||||| 
Db   454 TTTTGAGACGGAGTCTTGCTCTGTTGCCCAGGCTGGAGTGCAGTGGCGGCGATCTCAGCTC  395

Qy  1084 ACTGCAACCTCTGTCTCCCGGGTTCAAGCAATTCTCCATCTCAGCCTCCTGAGTAGCTG 1143
             ||||||| |||||                  ||||||| ||| |||||||||| |||||||
Db   394 ACTGCAAGCTCTGCCTCCCGGGTTCACGCCATTCTCCTGTCTCAGCCTCCTCAGTAGCTG  335

Qy  1144 GGA 1146
             |||
Db   334 GGA  332
```

RESULT 13
US-09-641-638-650/c
; Sequence 650, Application US/09641638
; Patent No. 6432648
; GENERAL INFORMATION:
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Chumakov, Ilya
; APPLICANT: Cohen, Annick
; TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
; TITLE OF INVENTION: GENES INVOLVED IN ARACHIDONIC ACID METABOLISM
; FILE REFERENCE: GENSET.051CP1
; CURRENT APPLICATION NUMBER: US/09/641,638
; CURRENT FILING DATE: 2000-08-16
; PRIOR APPLICATION NUMBER: US 09/502,330
; PRIOR FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: US 60/133,200
; PRIOR FILING DATE: 1999-05-07
; PRIOR APPLICATION NUMBER: US 09/275,267
; PRIOR FILING DATE: 1999-03-23
; PRIOR APPLICATION NUMBER: US 60/119,917
; PRIOR FILING DATE: 1999-02-12
; NUMBER OF SEQ ID NOS: 1304
; SOFTWARE: Patent.pm
; SEQ ID NO 650
; LENGTH: 1000
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: allele
; LOCATION: 501
; OTHER INFORMATION: 10-522-71 : polymorphic base A or G
; NAME/KEY: misc_binding
; LOCATION: 481..500
; OTHER INFORMATION: 10-522-71.mis1, potential
; NAME/KEY: misc_binding
; LOCATION: 502..521
; OTHER INFORMATION: 10-522-71.mis2, potential complement

; ORGANISM: Homo sapiens
US-09-489-847-22

Query Match   9.2%;   Score 106; DB 4;   Length 2876;
Best Local Similarity   85.5%;   Pred. No. 4.6e-23;
Matches 118; Conservative   0; Mismatches   20; Indels   0; Gaps   0;

Qy   1009 TCTTTTTCTTTTATTTTTGAGATGGAGTCTTGCTCTGTTGCCCAGCTGGAGTGCAGTG 1068
           | |||| ||||| |||| |||||||||||||||| ||| ||||||||||||||||||||
Db   2873 TTTTTTTTTTTTTTTTTTTTTTGAGATGGAATCTCACTCTGTGGCCAGGCTGGAGTGCAGTG 2814

Qy   1069 GTGTGATCTCAGCTTACTGCAACCTCTGTCTCCCGGGTTCAAGCAATTCTCCATCTCAG 1128
           |||||||||||||||||||||||||||||||||||| ||||||||||||||||| |||||
Db   2813 GAGCAATCTCAGCTCACTGCAGCCTCCGCCTCCCAGGTTCAAGCAATTCTCCTGCCTCAG 2754

Qy   1129 CCTCCTGAGTAGCTGGGA 1146
           ||||||||||||||||||
Db   2753 CCTCCTGAGTAGCTGGGA 2736

Search completed: December 1, 2003, 09:38:32
Job time : 95 secs

NAME/KEY: SITE
; LOCATION: (1284)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1287)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1378)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1912)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1913)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1935)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1947)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-489-847-111

Query Match   9.2%;   Score 106; DB 4;   Length 2871;
Best Local Similarity   85.5%;   Pred. No. 4.6e-23;
Matches 118; Conservative   0; Mismatches   20; Indels   0; Gaps   0;

Qy   1009 TCTTTTTCTTTTATTTTTGAGATGGAGTCTTGCTCTGTTGCCCAGCTGGAGTGCAGTG 1068
           | |||| ||||| |||| |||||||||||||||| ||| ||||||||||||||||||||
Db   2868 TTTTTTTTTTTTTTTTTTTTTTGAGATGGAATCTCACTCTGTGGCCAGGCTGGAGTGCAGTG 2809

Qy   1069 GTGTGATCTCAGCTTACTGCAACCTCTGTCTCCCGGGTTCAAGCAATTCTCCCATCTCAG 1128
           |||||||||||||||||||||||||||||||||||| ||||||||||||||||||| ||||
Db   2808 GAGCAATCTCAGCTCACTGCAGCCTCCGCCTCCCAGGTTCAAGCAATTCTCCTGCCTCAG 2749

Qy   1129 CCTCCTGAGTAGCTGGGA 1146
           ||||||||||||||||||
Db   2748 CCTCCTGAGTAGCTGGGA 2731

RESULT 15
US-09-489-847-22/c
; Sequence 22, Application US/09489847
; Patent No. 6476195
; GENERAL INFORMATION:
; APPLICANT: Rosen et al
; TITLE OF INVENTION: 98 Human Secreted Proteins
; FILE REFERENCE: PZ031P1
; CURRENT APPLICATION NUMBER: US/09/489,847
; CURRENT FILING DATE: 2000-01-24
; EARLIER APPLICATION NUMBER: PCT/US99/17130
; EARLIER FILING DATE: 1999-07-29
; EARLIER APPLICATION NUMBER: 60/094,657
; EARLIER FILING DATE: 1998-07-30
; EARLIER APPLICATION NUMBER: 60/095,486
; EARLIER FILING DATE: 1998-08-05
; EARLIER APPLICATION NUMBER: 60/096,319
; EARLIER FILING DATE: 1998-08-12
; EARLIER APPLICATION NUMBER: 60/095,454
; EARLIER FILING DATE: 1998-08-06
; EARLIER APPLICATION NUMBER: 60/095,455
; EARLIER FILING DATE: 1998-08-06
; NUMBER OF SEQ ID NOS: 376
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
; LENGTH: 2876
; TYPE: DNA

```
          GenCore version 5.1.6
     Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:    December 1, 2003, 07:30:33  ; Search time 1865 seconds
                                 (without alignments)
                           2026.279 Million cell updates/sec

Title:          US-10-049-742-22
Perfect score:  1147
Sequence:       1 gcctgggtcaagcagata........gcctcctgagtagctggat 1147

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       2190069 seqs, 1647345023 residues

Total number of hits satisfying chosen parameters:       4380138

Minimum DB seq length:  0
Maximum DB seq length:  2000000000

Post-processing:  Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

Database :      Published_Applications_NA:*
     1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
     2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
     3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
     4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
     5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
     6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
     7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
     8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
     9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
    10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
    11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
    12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
    13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
    14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
    15: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
    16: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
    17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

```
Result          %      Query
No.     Score   Match   Length  DB  Length  ID                          Description
-------------------------------------------------------------------------------------------------
c   1   1146    99.9    3090    12  3090    US-09-814-353-19835         Sequence 19835, A
c   2   1125    98.1    2945    14  2945    US-10-198-846-12415         Sequence 12415, A
    3   743     64.8    754     12  754     US-10-096-534-43            Sequence 43, Appl
    4   743     64.8    754     14  754     US-10-097-340-182           Sequence 182, App
    5   453     39.5    7453    11  7453    US-09-764-891-8949          Sequence 8949, Ap
c   6   293.2   25.6    443     12  443     US-09-814-353-14171         Sequence 14171, A
c   7   257     22.4    259     12  259     US-09-814-353-1424          Sequence 1424, Ap
c   8   257     22.4    259     12  259     US-09-814-353-7786          Sequence 7786, Ap
c   9   251.4   21.9    12572   14  12572   US-10-228-406A-10           Sequence 10, Appl
    10  251.4   21.9    14078   12  14078   US-10-134-288-1             Sequence 1, Appli
    11  251.4   21.9    16713   14  16713   US-10-228-406A-9            Sequence 9, Appli
    12  234.6   20.5    339     10  339     US-09-867-701-8500          Sequence 8500, Ap
c   13  137.6   12.0    510     12  510     US-09-814-353-4106          Sequence 4106, Ap
c   14  137.6   12.0    510     12  510     US-09-814-353-10412         Sequence 10412, A
c   15  137     11.9    598     12  598     US-09-814-353-16796         Sequence 16796, A
c   16  137     11.9    621     14  621     US-10-198-846-9074          Sequence 9074, Ap
```

```
c   17  116.8   10.2    28001   13  US-10-193-295-3           Sequence 3, Appli
    18  115     10.0    403     12  US-10-027-632-279852      Sequence 279852,
    19  115     10.0    403     13  US-10-027-632-279852      Sequence 279852,
    20  114     9.9     48763   14  US-10-282-048-3           Sequence 3, Appli
c   21  113.6   9.9     498     10  US-09-920-300A-1487       Sequence 1487, Ap
    22  113.6   9.9     498     12  US-10-099-926-1487        Sequence 1487, Ap
    23  113.6   9.9     498     13  US-10-033-528-1487        Sequence 1487, Ap
    24  112.8   9.8     2147    10  US-09-764-841-1738        Sequence 1738, Ap
    25  112.8   9.8     2147    11  US-10-092-154-1738        Sequence 1738, Ap
    26  112.8   9.8     32174   9   US-09-908-711-158         Sequence 158, App
    27  122.8   9.8     32174   9   US-09-764-860-1134        Sequence 1134, Ap
    28  112.8   9.8     32174   10  US-09-764-877-2645        Sequence 2645, Ap
    29  112.8   9.8     32174   10  US-09-860-670-232         Sequence 232, App
    30  112.8   9.8     32174   10  US-09-764-904-90          Sequence 90, Appl
    31  112.8   9.8     32174   11  US-09-764-891-6480        Sequence 6480, Ap
    32  112.8   9.8     32174   11  US-09-764-891-10135       Sequence 10135, A
    33  112.8   9.8     32174   11  US-09-764-891-10179       Sequence 10179, A
    34  112.8   9.8     32174   13  US-10-212-872-1134        Sequence 1134, Ap
    35  112.8   9.8     32174   14  US-10-091-548-90          Sequence 90, Appl
    36  112.8   9.8     32174   14  US-10-074-095-1134        Sequence 1134, Ap
    37  112.4   9.8     174493  10  US-09-804-471A-3          Sequence 3, Appli
    38  112.4   9.8     174493  14  US-10-238-709-3           Sequence 3, Appli
c   39  112.2   9.8     12149   9   US-09-764-869-2258        Sequence 2258, Ap
c   40  112.2   9.8     12149   14  US-10-091-504-2258        Sequence 2258, Ap
    41  112.2   9.8     106344  11  US-09-910-185-10          Sequence 10, Appl
    42  112     9.8     689     12  US-10-027-632-135995      Sequence 135995,
    43  112     9.8     689     13  US-10-027-632-135995      Sequence 135995,
    44  112     9.8     1536    12  US-10-027-632-252944      Sequence 252944,
    45  112     9.8     1536    13  US-10-027-632-252944      Sequence 252944,
```

ALIGNMENTS

Query Match   99.9%; Score 1146; DB 12; Length 3090;

Db   647 TATTTTTGAGATGAGTCTTGCTCTGTTGCCCAGGCTGGAGTGGTGTGATCTCAG 588
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 1081 CTTACTGCAACCTCGTCTCCCGGGTTCAAGCAATTCTCCCATTCCAGCCTCCGAGTAG 1140
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   587 CTTACTGCAACCTCGTCTCCCGGGTTCAAGCAATTCTCCCATTCCAGCCTCCGAGTAG 528

Qy 1141 CTGGGAT 1147
         || ||||
Db   527 CTNGGAT 521

RESULT 2
US-10-198-846-12415/c
; Sequence 12415, Application US/10198846
; Publication No. US20030099974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO: 12415
; LENGTH: 2945
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-198-846-12415

Query Match   98.1%; Score 1125; DB 14; Length 2945;
Best Local Similarity   99.8%; Pred. No. 0;
Matches 1147; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

Qy   1 GCCTTGGGTCAAGCAGCAGAATATTAATAGGCAGGGAATGCACCTGTAGCTACTGGGCGCTA 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1673 GCCTTGGGTCAAGCAGCAGAATATTAATAGGCAGGGAATGCACCTGTAGCTAGTGGGCGCTA 1614

Qy   61 CTGCCAGCCTGAAGAGGAAGTGGCTCGACTCTTGACCATGGCTGGGGTTCCTGAGGATGA 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1613 CTGCCAGCCTGAAGAGGAAGTGGCTCGACTCTTGACCATGGCTGGGGTTCCTGAGGATGA 1554

Qy   121 GCTAAACCCTTCCATGTACTGGGGGTTGAGCGCCACAGCATCAGATGTTGAACTGAAGAA 180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1553 GCTAAACCCTTCCATGTACTGGGGGTTGAGCGCCACAGCATCAGATGTTGAACTGAAGAA 1494

Qy   181 GGCCTATAGACAGCTGGCAGTGATGGTTCATCCTGACAAAAATCATCATCCCGGGCTGA 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1493 GGCCTATAGACAGCTGGCAGTGATGGTTCATCCTGACAAAAATCATCATCCCGGGCTGA 1434

Qy   241 GGAGGCCTTCAAGGTTTGCGAGGTGGTGGGGTTGAGGCCAGCAATGGTTGCAGAAGCGAAA 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1433 GGAGGCCTTCAAGGTTTGCGAGGTGGTGGGGTTGAGGCCAGCAATGGTTGCAGAAGCGAAA 1374

Qy   301 GGAGTATGAGATGAAACGAATGGCAGAGAATGAGCTCAGTAGCTGAGCCGGTCAGTAGAGTTTCT 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1373 GGAGTATGAGATGAAACGAATGGCAGAGAATGAGCTCAGTAGCTGAGCCGGTCAGTAGAGTTTCT 1314

Qy   361 GTCCAAGCTGCAAGATGACCTCAAGGAGCAATGAATACATGATGTGTAGCCGATGCCA 420
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1313 GTCCAAGCTGCAAGATGACCTCAAGGAGCAATGAATACATGATGTGTAGCCGATGCCA 1254

Qy   421 AGGAAAGCATAGGAGGTTTGAAATGGACCCGGAACCTAAGAGTGCCAGATACTGTCTGA 480
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1253 AGGAAAGCATAGGAGGTTTGAAATGGACCCGGAACCTAAGAGTGCCAGATACTGTCTGA 1194

Qy   481 GTGTAATAGGCTGCATCCTGCTGAGGAAGGAGACTTTTGGGCAGAGTCAAGCATGTTGGG 540

Best Local Similarity   99.9%; Pred. No. 0;
Matches 1146; Conservative 0; Mismatches 0; Indels 1; Gaps 0;

Qy   1 GCCTTGGGTCAAGCAGCAGAATATTAATAGGCAGGGAATGCACCTGTAGCTACTGGGCGCTA 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1667 GCCTTGGGTCAAGCAGCAGAATATTAATAGGCAGGGAATGCACCTGTAGCTAGTGGGCGCTA 1608

Qy   61 CTGCCAGCCTGAAGAGGAAGTGGCTCGACTCTTGACCATGGCTGGGGTTCCTGAGGATGA 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1607 CTGCCAGCCTGAAGAGGAAGTGGCTCGACTCTTGACCATGGCTGGGGTTCCTGAGGATGA 1548

Qy   121 GCTAAACCCTTCCATGTACTGGGGGTTGAGCGCCACAGCATCAGATGTTGAACTGAAGAA 180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1547 GCTAAACCCTTCCATGTACTGGGGGTTGAGCGCCACAGCATCAGATGTTGAACTGAAGAA 1488

Qy   181 GGCCTATAGACAGCTGGCAGTGATGGTTCATCCTGACAAAAATCATCATCCCGGGCTGA 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1487 GGCCTATAGACAGCTGGCAGTGATGGTTCATCCTGACAAAAATCATCATCCCGGGCTGA 1428

Qy   241 GGAGGCCTTCAAGGTTTGCGAGCCGTTCATCCTGGACACATTGTCAGCAATGCTGAAAAGCGAAA 300
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1427 GGAGGCCTTCAAGGTTTGCGAGCCGTTCATCCTGGACACATTGTCAGCAATGCTGAAAAGCGAAA 1368

Qy   301 GGAGTATGAGATGAAACGAATGGCAGAGAATGAGCTCAGCCGGTCAGTAAATGAGTTTCT 360
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1367 GGAGTATGAGATGAAACGAATGGCAGAGAATGAGCTCAGCCGGTCAGTAAATGAGTTTCT 1308

Qy   361 GTCCAAGCTGCAAGATGACCTCAAGGAGCAATGAATACATGATGTGTAGCCGATGCCA 420
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1307 GTCCAAGCTGCAAGATGACCTCAAGGAGCAATGAATACATGATGTGTAGCCGATGCCA 1248

Qy   421 AGGAAAGCATAGGAGGTTTGAAATGGACCCGGAACCTAAGAGTGCCAGATACTGTCTGA 480
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1247 AGGAAAGCATAGGAGGTTTGAAATGGACCCGGAACCTAAGAGTGCCAGATACTGTCTGA 1188

Qy   481 GTGTAATAGGCTGCATCCTGCTGAGGAAGGAGACTTTTGGGCAGAGTCAAGCATGTTGGG 540
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1187 GTGTAATAGGCTGCATCCTGCTGAGGAAGGAGACTTTTGGGCAGAGTCAAGCATGTTGGG 1128

Qy   541 CCTCAAGATCACCACTTTGCACTGATGATGGAAGAGGTGTATGACAATCACAGAGTGGGC 600
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1127 CCTCAAGATCACCACTTTGCACTGATGATGGAAGAGGTGTATGACAATCACAGAGTGGGC 1068

Qy   601 TGGATGCCAGCGTGTAGGTATCTCCCCAGATACCCACAGAGTCCCCATCACATCTCATT 660
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1067 TGGATGCCAGCGTGTAGGTATCTCCCCAGATACCCACAGAGTCCCCATCACATCTCATT 1008

Qy   661 TGGTTCTCGGATTCTTCTGAGTCGGATCTTTCAAGTACCCCAGGGCAGATGCCCAATGG 720
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1007 TGGTTCTCGGATTCTTCTGAGTCGGATCTTTCAAGTACCCCAGGGCAGATGCCCCTCCTGC 948

Qy   721 TGATCTTCAGGATTCTTCTGAGTCGGATCTTTCAAGTACCCCAGGGCAGATGCCCAATGG 780
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   947 TGATCTTCAGGATTCTTCTGAGTCGGATCTTTCAAGTACCCCAGGGCAGATGCCCAATGG 888

Qy   781 GAACTTCTTTGCAGTCCTCAGCAGCCAAACCTAAGCGCGCGGAAGAAAGTGAGGAGGCCTTCCA 840
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   887 GAACTTCTTTGCAGTCCTCAGCAGCCAAACCTAAGCGCGCGGAAGAAAGTGAGGAGGCCTTCCA 828

Qy   841 CACAGTACCCAAGGGAGAAGCCAAAACCTAAGCGGCGGAAAAGGTGAGGAGAAGGGGCCCTTCCA 900
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   827 CACAGTACCCAAGGGAGAAGCCAAAACCTAAGCGGCGGAAAAGGTGAGGAGAAGGGGCCCTTCCA 768

Qy   901 ACGTTGATGCCCCTTTTCTCTTTTCCTCAAATCAATGTCAAAAGGGGTCTGTAGCAC 960
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   767 ACGTTGATGCCCCTTTTCTTTTTCCTCAAATCAATGTCAAAAGGGGTCTGTAGCAC 708

Qy   961 AGGATGGAGTTTGATTTATCCTCTCCCCAAACACCTAGGAACTGAATCTTTTTTCTTT 1020
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   707 AGGATGGAGTTTGATTTTATCCTCTCCCCAAACACCTAGGAACTGAATCTTTTTTCTTT 648

Qy 1021 TATTTTTGAGATGAGTCTTGCTCTGTTGCCCAGGCTGGAGTGGTGTGATCTCAG 1080

Matches 754; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

```
Qy   301 GGAGTATGAGATGAAACGAATGCAGAGAATGAGCTGAGCCGGTCAGTCAGTAAATGAGTTCT 360
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 GGAGTATGAGATGAAACGAATGCAGAGAATGAGCTGAGCCGGTCAGTCAGTAAATGAGTTCT  60

Qy   361 GTCCAAGCTGCAAGATGACCTCAAGAGGAGGCAATGAATGAATACTATGTGTAGCCGATGCCA 420
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 GTCCAAGCTGCAAGATGACCTCAAGAGGAGGCAATGAATGAATACTATGTGTAGCCGATGCCA 120

Qy   421 AGGAAAGCATAGGAGGTTGAAATGAGGACCGGGAACCTAAGAGTGCCAGATACTGTGCTGA 480
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 AGGAAAGCATAGGAGGTTGAAATGAGGACCGGGAACCTAAGAGTGCCAGATACTGTGCTGA 180

Qy   481 GTGTAATAGGCTGCATCCTGCTGAGGAAGGAGACTTTTGGGCAGAGTCAAGCATGTTGGG 540
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 GTGTAATAGGCTGCATCCTGCTGAGGAAGGAGACTTTTGGGCAGAGTCAAGCATGTTGGG 240

Qy   541 CCTCAAGATCACCTACTTTGCACTGATGGATGGAAAGGTGTATGACATCACAGAGTGGGC 600
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 CCTCAAGATCACCTACTTTGCACTGATGGATGGAAAGGTGTATGACATCACAGAGTGGGC 300

Qy   601 TGGATGCCAGCGTGTAGGTATCTCCCAGATACCCACAGAGTCCCTATCACATCTCATT 660
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 TGGATGCCAGCGTGTAGGTATCTCCCAGATACCCACAGAGTCCCTATCACATCTCATT 360

Qy   661 TGGTTCTCGGATTCCAGGCACCACCAGAGAGCGGCGGAGATCTTCAAGTACCCCAATGG 720
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 TGGTTCTCGGATTCCAGGCACCACCAGAGAGCGGCGGAGATCTTCAAGTACCCCAATGG 420

Qy   721 TGATCTTCAGGATTTCTTGAGTCGGATCTTCGCCCTGAGCCCGGCCAGTGCCCAACAG 780
         |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db   421 TGATCTTCAGGATTTCTTGAGTCGGATCTTCGCCCTGAGCCCGGCCAGTG-CCAATGG 479

Qy   781 GAACTTCTTTGCAGCTCCTCAGCCTGCCCTGAGCCCTCAAGCCCAACAG 840
         |||||||||||||||||||||||||||||||||||||||||||||||||
Db   480 GAACTTCTTTGCAGCTCCTCAGCCTGCCCTGAGCCCTCAAGCCCAACAG 539

Qy   841 CACAGTACCCAAGGGAGAAGCCAAACCTAAGGCGGGAAGAAAGTGAGGAGCGGGCCCTTCCA 900
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   540 CACAGTACCCAAGGGAGAAGCCAAACCTAAGGCGGGAAGAAAGTGAGGAGCGGGCCCTTCCA 599

Qy   901 ACGTTGATGCCCCTTCTCTTTGATTATCCCTCCCCCAACACCTAGGAACTCAATGTCAGGGAGTCAGGGCAC 960
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   600 ACGTTGATGCCCCTTCTCTTTGATTATCCCTCCCCCAACACCTAGGAACTCAATGTCAGGGAGTCAGGGCAC 659

Qy   961 AGGATGGAGTTTGATTTATCCCTCCTCCCCCCAACACCTAGGAACTGAATCTTTTTCTTTT 1020
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   660 AGGATGGAGTTTGATTTATCCCTCCTCCCCCCAACACCTAGGAACTGAATCTTTTTCTTTT 719

Qy  1021 TATTTTTTGAGATGGAGTCTTGCTCTGTTGCCCAG 1055
         |||||||||||||||||||||||||||||||||||
Db   720 TATTTTTTGAGATGGAGTCTTGCTCTGTTGCCCAG 754
```

RESULT 3
US-10-096-534-43
; Sequence 43, Application US/10096534
; Publication No. US20030166887A1
; GENERAL INFORMATION:
;   APPLICANT: The Brigham and Women's Hospital, Inc.
;   APPLICANT: Yates, Karen
;   APPLICANT: Mizuno, Shuichi
;   APPLICANT: Glowacki, Julie
;   TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF SKELETAL DEGENERATION CONDITIONS
;   FILE REFERENCE: B0801/7244/KA/ERP
;   CURRENT APPLICATION NUMBER: US/10/096,534
;   CURRENT FILING DATE: 2002-03-12
;   PRIOR APPLICATION NUMBER: US 60/274,980
;   PRIOR FILING DATE: 2001-03-12
;   NUMBER OF SEQ ID NOS: 79
;   SOFTWARE: PatentIn version 3.0
; SEQ ID NO 43
;   LENGTH: 754
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-096-534-43

```
Db  1193 GTGTAATAGGCTGCATCCTGCTGAGGAAGGAGACTTTTGGGCAGAGTCAAGCATGTTGGG 1134
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   541 CCTCAAGATCACCTACTTTGCACTGATGGATGGAAAGGTGTATGACATCACAGAGTGGGC 600

Db  1133 CCTCAAGATCACCTACTTTGCACTGATGGATGGAAAGGTGTATGACATCACAGAGTGGGC 1074
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   601 TGGATGCCAGCGTGTAGGTATCTCCCAGATACCCACAGAGTCCCTATCACATCTCATT 660

Db  1073 TGGATGCCAGCGTGTAGGTATCTCCCAGATACCCACAGAGTCCCTATCACATCTCATT 1014
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   661 TGGTTCTCGGATTCCAGGCACCACCAGAGAGCGGCGGAGATCTTCAAGTACCCCAATGG 720

Db  1013 TGGTTCTCGGATTCCAGGCACCACCAGAGAGCGGCGGAGATCTTCAAGTACCCCAATGG 954
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   721 TGATCTTCAGGATTTCTTGAGTCGGATCTTCGCCCTGAGCCCGGCCAGTGCCCAATGG 780

Db   953 TGATCTTCAGGATTTCTTGAGTCGGATCTTCGCCCTGAGCCCGGCCAGTGCCCAATGG 894
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   781 GAACTTCTTTGCAGCTCCTCAGCCTGCCCTGAGCCCTCAAGCCCAACAG 840

Db   893 GAACTTCTTTGCAGCTCCTCAGCCTGCCCTGAGCCCTCAAGCCCAACAG 834
         |||||||||||||||||||||||||||||||||||||||||||||||||
Qy   841 CACAGTACCCAAGGGAGAAGCCAAACCTAAGGCGGGAAGAAAGTGAGGAGCGGGCCCTTCCA 900

Db   833 CACAGTACCCAAGGGAGAAGCCAAACCTAAGGCGGGAAGAAAGTGAGGAGCGGGCCCTTCCA 774
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   901 ACGTTGATGCCCCTTCTCTTTGATTATCCCTCCCCCAACACCTAGGAGTCAGGGCAC 960

Db   773 ACGTTGATGCCCCTTCTCTTTGATTATCCCTCCCCCAACACCTAGGAGTCAGGGCAC 714
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   961 AGGATGGAGTTTGATTTATCCCTCCGGG-TTCAAGCAATTCTCCCATCTCAGCCTCCTGAGT 1020

Db   713 AGGATGGAGTTTGATTTATCCCGGGCTTCAAGCAATTCTCCCATCTCAGCCTCCTGAGT 654
         |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Qy  1021 TATTTTTTGAGATGGAGTCTTGCTCT-GTTGCCCAGGCTGGAGTGCAGTGGTGTGATCTCA 1079

Db   594 GCTTACTGCAACCTCTGTCTCCCGGGGCTTCAAGCAATTCTCCCATCTCAGCCTCCTGAGT 1138

Db   534 GCTTACTGCAACCTCTGTCTCCCGGGGCTTCAAGCAATTCTCCCATCTCAGCCTCCTGAGT 594

Qy  1139 AGCTGGGAT 1147
         |||||||||
Db   533 AGCTGGGAT 525
```

Query Match   64.8%;   Score 743; DB 12; Length 754;
Best Local Similarity   99.9%;   Pred. No. 1.7e-232;

RESULT 4
US-10-097-340-182
; Sequence 182, Application US/10097340
; Publication No. US20030087250A1
; GENERAL INFORMATION:
;   APPLICANT: John MONAHAN
;   APPLICANT: Marjula GANNAVARAPU
;   APPLICANT: Sebastian HOERSCH
;   APPLICANT: Shubhangi KAMATKAR
;   APPLICANT: Steve G. KOVATS
;   APPLICANT: Rachel E. MEYERS
;   APPLICANT: Michael MORRISEY
;   APPLICANT: Peter OLANDT
;   APPLICANT: Ami SEN
;   APPLICANT: Peter VEIBY
;   APPLICANT: Gordon B. MILLS
;   APPLICANT: Robert C. BAST, Jr.
;   APPLICANT: Karen LU

```
; APPLICANT: Rosemarie SCHMANDT
; APPLICANT: Xumei ZHAO
; APPLICANT: Karen GLATT
; TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,
; TITLE OF INVENTION: Assessment, Prevention, and Therapy of Ovarian Cancer
; FILE REFERENCE: MRI-030
; CURRENT APPLICATION NUMBER: US/10/097,340
; CURRENT FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: 60/276,025
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/325,149
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/276,026
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/324,967
; PRIOR FILING DATE: 2001/09/26
; PRIOR APPLICATION NUMBER: 60/311,732
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/325,102
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/323,580
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 363
; SEQ ID NO 182
; LENGTH: 754
; TYPE: DNA
; ORGANISM: Homo sapiens
; SOFTWARE: FastSEQ for Windows Version 4.0
US-10-097-340-182

Query Match             64.8%;  Score 743;  DB 14;  Length 754;
Best Local Similarity   99.9%;  Pred. No. 1.7e-232;
Matches   754;  Conservative   0;  Mismatches   0;  Indels   1;  Gaps   1;

Qy      301 GGAGTATGAGATGAAACGATGCAGAGAATGAGCTGAGCCGGTCAGTGAGTTTCT 360
                |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 GGAGTATGAGATGAAACGATGCAGAGAATGAGCTGAGCCGGTCAGTGAGTTTCT 60

Qy      361 GTCCAAGCTGCAAGATGACCTCAAGGAGGCAATGAATACTATGATGTGTAGCCGATGCCA 420
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 GTCCAAGCTGCAAGATGACCTCAAGGAGGCAATGAATACTATGATGTGTAGCCGATGCCA 120

Qy      421 AGGAAAGCATAGGAGTTGAAATGGACCGGGAACCTAAGAGTGCCAGATACTGTGCTGA 480
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 AGGAAAGCATAGGAGTTGAAATGGACCGGGAACCTAAGAGTGCCAGATACTGTGCTGA 180

Qy      481 GTGTAATAGGCTGCATCCTGCTGAGGAGGAGGACTTTTGGGCAGAGTCAAGCATGTTGGG 540
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      181 GTGTAATAGGCTGCATCCTGCTGAGGAGGAGGACTTTTGGGCAGAGTCAAGCATGTTGGG 240

Qy      541 CCTCAAGATCACCTACTTTGCACTGATGGAAAGGTGTATGACATCACAGAGTGGGC 600
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      241 CCTCAAGATCACCTACTTTGCACTGATGGAAAGGTGTATGACATCACAGAGTGGGC 300

Qy      601 TGGATGCCAGCGTGTGGTAGGTATCTCCCAGAGTCCCCTATCACATCTCATT 660
                |||||||||||||||||||||||||||||||||||||||||||||||||||
Db      301 TGGATGCCAGCGTGTGGTAGGTATCTCCCAGAGTCCCCTATCACATCTCATT 360

Qy      661 TGGTTCTCGGATTCCAGGCACCAGAGAGCGGGCAGAGAGCCCCTCCTGC 720
                ||||||||||||||||||||||||||||||||||||||||||||||||
Db      361 TGGTTCTCGGATTCCAGGCACCAGAGAGCGGGCAGAGAGCCCCTCCTGC 420

Qy      721 TGATCTTCAGATTCTTGAGTCGATCTTCAAGCCTGCCCCTGAGCCTTAAGCCAACAG 780
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      421 TGATCTTCAGATTCTTGAGTCGATCTTCAAGCCTGCCCCTGAGCCTTAAGCCAACAG 480

Qy      781 GAACTTCTTGCAGCTCCTCAGCCTCCTGCCCGCCTGCCGCTCAGCCCTTAAGCCCAACAG 840
                ||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db      480 GAACTTCTTGCAGCTCCTCAGCCTCCTGCCCGCCTGCCGCTCAGCCCTTAAGCCCAACAG 539

Qy      841 CACAGTACCCAAGGGAGAAGCCAAACCTAAGCCGGCGAAGAAAGTGAGGAGCCCTTCCA 900
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      540 CACAGTACCCAAGGGAGAAGCCAAACCTAAGCCGGCGAAGAAAGTGAGGAGCCCTTCCA 599

Qy      901 ACGTTGATGCCCCTCTTCCTCAAATCAATGTCAGGAGTCAAAAGGGCTGTAGCAC 960
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      600 ACGTTGATGCCCCTCTTCCTCAAATCAATGTCAGGAGTCAAAAGGGCTGTAGCAC 659

Qy      961 AGGATGGAGTTGATTTATCCTCCTCCCCAACACTAGGAACTGAATCTTTTCTTTT 1020
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      660 AGGATGGAGTTGATTTATCCTCCTCCCCAACACTAGGAACTGAATCTTTTCTTTT 719

Qy     1021 TATTTTTGAGATGGAGTCTTGCTCTGTTGCCCAG 1055
                ||||||||||||||||||||||||||||||||||
Db      720 TATTTTTGAGATGGAGTCTTGCTCTGTTGCCCAG 754


RESULT 5
US-09-764-891-8949
; Sequence 8949, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8949
; LENGTH: 7453
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (350)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-891-8949

Query Match             39.5%;  Score 453;  DB 11;  Length 7453;
Best Local Similarity  100.0%;  Pred. No. 6e-137;
Matches   453;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy      695 AGAGCCACCCCAGATGCCCCTCCTGCTGCTCATCTTCAGGATTCTTGAGTCGGATCTTTCAA 754
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3315 AGAGCCACCCCAGATGCCCCTCCTGCTGCTCATCTTCAGGATTCTTGAGTCGGATCTTTCAA 3374

Qy      755 GTACCCCAGGGCAGATGCCAATGCCAGGGAACTTCTTGCAGCTCCTCAGCGTCCTGCCCTGA 814
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3375 GTACCCCAGGGCAGATGCCAATGCCAGGGAACTTCTTGCAGCTCCTCAGCGTCCTGCCCTGA 3434

Qy      815 GCCCGTGCAGCCTCAAGCCAACAGCAGCAAACCTAAGCGG 874
                ||||||||||||||||||||||||||||||||||||||||
Db     3435 GCCCGTGCAGCCTCAAGCCAACAGCAGCAAACCTAAGCGG 3494

Qy      875 CGGAAGAAAGTGAGGAGGCCCTTCCAAGTTGATGCCCCTTCCTCTTTCTTCCTCACCCCCAAC 934
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3495 CGGAAGAAAGTGAGGAGGCCCTTCCAAGTTGATGCCCCTTGATCCCCTTCTTCCTCACCCCCAAC 3554

Qy      935 TCAGGGAGTCAAAAGGGCTGTAGCACAGGATGGAGTTTGATTTATCCCTCCCCCAAC 994
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3555 TCAGGGAGTCAAAAGGGCTGTAGCACAGGATGGAGTTTGATTTATCCCTCCCCCAAC 3614

Qy      995 ACCTAGGAACTGAATCTTTTTCTTTTTTATTTTTTGAGATGGAGTCTTGTTGCCCA 1054
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3615 ACCTAGGAACTGAATCTTTTTCTTTTTTATTTTTTGAGATGGAGTCTTGTTGCCCA 3674

Qy     1055 GCTGGAGTGCAGTGGTGTGATCTCAGCTTACTGCAACCTCGTCTCCCGGTTCAAGCAA 1114
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3675 GCTGGAGTGCAGTGGTGTGATCTCAGCTTACTGCAACCTCGTCTCCCGGTTCAAGCAA 3734

Qy     1115 TTCTCCCATCTCAGCCTCCTGAGTAGCTGGGAT 1147
                |||||||||||||||||||||||||||||||||
Db     3735 TTCTCCCATCTCAGCCTCCTGAGTAGCTGGGAT 3767
```

RESULT 6
US-09-814-353-14171/c
; Sequence 14171, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14171
; LENGTH: 443
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-814-353-14171

Query Match   25.6%; Score 293.2; DB 12; Length 443;
Best Local Similarity   98.7%; Pred. No. 3.2e-85;
Matches 306; Conservative   0; Mismatches   3; Indels   1; Gaps   1;

Qy   133 CCATGTACTGGGGGTTGAGGCCACAGCATCAGA-TGTTGAACTGAAGAAGGCCTATAGAC 191
           ||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db   426 CAAGGTACTGGGGGTTAAGGCCACAGCATCAGATTGTTGAACTGAAGAAGGCCTATAGAC 367

Qy   192 AGCTGGCAGTGATGGTTCATCCTGACAAAATCATCATCCCCGGGCTGAGGAGGCCTTCA 251
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   366 AGCTGGCAGTGATGGTTCATCCTGACAAAATCATCATCCCCGGGCTGAGGAGGCCTTCA 307

Qy   252 AGGTTTTGCGAGCAGCTTGGGACATTGTCAGCAATGCTGAAAAGCGAAAGGAGTATGAGA 311
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   306 AGGTTTTGCGAGCAGCTTGGGACATTGTCAGCAATGCTGAAAAGCGAAAGGAGTATGAGA 247

Qy   312 TGAAACGAATGGCAGAGAATGAGCTGAGCCGGTCAGTAAATGAGTTTCTGTCCAAGCTGC 371
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   246 TGAAACGAATGGCAGAGAATGAGCTGAGCCGGTCAGTAAATGAGTTTCTGTCCAAGCTGC 187

Qy   372 AAGATGACCTCAAGGAGGCAATGAATACTATGATGTGTAGCCGATGCCAAGGAAAGCATA 431
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   186 AAGATGACCTCAAGGAGGCAATGAATACTATGATGTGTAGCCGATGCCAAGGAAAGCATA 127

Qy   432 GGAGGTTTGA 441
           ||||||||||
Db   126 GGAGGTTTGA 117

RESULT 7
US-09-814-353-1424/c
; Sequence 1424, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR

; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1424
; LENGTH: 259
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1, 2
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-1424

Query Match   22.4%; Score 257; DB 12; Length 259;
Best Local Similarity 100.0%; Pred. No. 1.7e-73;
Matches 257; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   184 CTATAGACAGCTGCAGTGATGGTTCATCCTGACAAAATCATCATCCCCGGGCTGAGGA 243
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   259 CTATAGACAGCTGCAGTGATGGTTCATCCTGACAAAATCATCATCCCCGGGCTGAGGA 200

Qy   244 GGCCTTCAAGGTTTTGCGAGCAGCTTGGGACATTGTCAGCAATGCTGAAAAGCGAAAGGA 303
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   199 GGCCTTCAAGGTTTTGCGAGCAGCTTGGGACATTGTCAGCAATGCTGAAAAGCGAAAGGA 140

Qy   304 GTATGAGATGAAACGAATGGCAGAGAATGAGCTGAGCCGGTCAGTAAATGAGTTTCTGTC 363
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   139 GTATGAGATGAAACGAATGGCAGAGAATGAGCTGAGCCGGTCAGTAAATGAGTTTCTGTC 80

Qy   364 CAAGCTGCAAGATGACCTCAAGGAGGCAATGAATACTATGATGTGTAGCCGATGCCAAGG 423
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   79 CAAGCTGCAAGATGACCTCAAGGAGGCAATGAATACTATGATGTGTAGCCGATGCCAAGG 20

Qy   424 AAAGCATAGGAGGTTTG 440
           |||||||||||||||||
Db   19 AAAGCATAGGAGGTTTG 3

RESULT 8
US-09-814-353-7786/c
; Sequence 7786, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940

```
;   PRIOR FILING DATE: 2000-06-15
;   PRIOR APPLICATION NUMBER: US 60/216,820
;   PRIOR FILING DATE: 2000-07-07
;   PRIOR APPLICATION NUMBER: US 60/220,661
;   PRIOR FILING DATE: 2000-07-25
;   PRIOR APPLICATION NUMBER: US 60/257,672
;   PRIOR FILING DATE: 2000-12-21
;   NUMBER OF SEQ ID NOS: 22037
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 7786
;   LENGTH: 259
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 1, 2
;   OTHER INFORMATION: n = A,T,C or G
US-09-814-353-7786

Query Match          22.4%;   Score 257;  DB 12;   Length 259;
Best Local Similarity  100.0%;  Pred. No. 1.7e-73;
Matches  257;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy   184  CTATAGACAGCTGGCAGTGATGGTTCATCCTGACAAAAATCATCATCCCCGGCTGAGGA  243
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   259  CTATAGACAGCTGGCAGTGATGGTTCATCCTGACAAAAATCATCATCCCCGGCTGAGGA  200

Qy   244  GGCCTTCAAGGTTTGCGAGCAGCTTGGGACATTGTCAGCAATGCTGAAAAGCGAAAGGA  303
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   199  GGCCTTCAAGGTTTGCGAGCAGCTTGGGACATTGTCAGCAATGCTGAAAAGCGAAAGGA  140

Qy   304  GTATGAGATGAAAACGAATGGCAGAGAATGAGCTGAGCCGGTCAGTAAATGAGTTTCTGTC  363
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   139  GTATGAGATGAAAACGAATGGCAGAGAATGAGCTGAGCCGGTCAGTAAATGAGTTTCTGTC  80

Qy   364  CAAGCTGCAAGATGACCTCAAGGAGGCAATGAATACTATGATGTGTAGCCGATGCCAAGG  423
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   79   CAAGCTGCAAGATGACCTCAAGGAGGCAATGAATACTATGATGTGTAGCCGATGCCAAGG  20

Qy   424  AAAGCATAGGAGGTTTG  440
          |||||||||||||||||
Db   19   AAAGCATAGGAGGTTTG  3

RESULT 9
US-10-228-406A-10
;   Sequence 10, Application US/10228406A
;   Publication No. US20030104612A1
;   GENERAL INFORMATION:
;   APPLICANT: Cao, Xuemei
;   APPLICANT: Zybarth, Gabriele
;   TITLE OF INVENTION: GENERATION OF TYPE I/TYPE II HYBRID FORM OF BOVINE
;   TITLE OF INVENTION: VIRAL DIARRHEA VIRUS FOR USE AS VACCINE
;   FILE REFERENCE: PC11051A
;   CURRENT APPLICATION NUMBER: US/10/228,406A
;   CURRENT FILING DATE: 2002-08-27
;   NUMBER OF SEQ ID NOS: 12
;   SOFTWARE: PatentIn Ver. 2.1
;   SEQ ID NO 10
;   LENGTH: 12572
;   TYPE: RNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Hybrid BVD virus NADL890
US-10-228-406A-10

Query Match          21.9%;   Score 251.4;  DB 14;   Length 12572;
Best Local Similarity  72.2%;  Pred. No. 9.2e-71;
Matches  200;  Conservative  61;  Mismatches  16;  Indels  0;  Gaps  0;

Qy   404  ATGTGTAGCCGATGCCAAGGAAAGCATAGGAGGTTTGAAATGGACCGGGAACCTAAGAGT  463
          |:|:|.|||||:|||| ||||||||||||||:|| ||:|||||||||||||||:||||||
Db   4988 AUGUGCAGCCGAUGCCAGGGAAAGCAUAGGAGGUUUGAAAUGGACCGGGAACCUAAGAGU  5047
```

```
Qy   464  GCCAGATACTGTCTGAGTGTAATAGGCTGCATCCTGCTGAGGAAGGAGACTTTTGGGCA  523
          |||:|.|||||:||:|:|:|||:||||:||||||||:|||||||||||||:|||:||||
Db   5048 GCCAGAUACUGUCUGAGUGUAAUAGGCUGCAUCCUGCUGAGGAAGGUGACUUUUGGGCA  5107

Qy   524  GAGTCAAGCATGTTGGGCCTCAAGATCACCTACTTTGCACTGATGATGATGAAAGGTGAT  583
          |||:.| ||||||||||||||||||:|.:||||::||::||||:|:||:||:|||:||:
Db   5108 GAGUCGAGCAUGUUGGGCCUCAAAUCACCUACUUUGCGCUGAUGAUGGAAAGGUGAU  5167

Qy   584  GACATCACAGAGTGGGCTGGATGCCAGCGTGTAGGTATCTCCCAGATACCCACAGAGTC  643
          ||.|:.|||||:| |||:|||||||||||||||.||||:|||||||||:|||||||:|
Db   5168 GAUAUCACAGAGUGGGCUGGAUGCCAGCGUGUGGGAAUCCCCAGAUACCCACAGAGUC  5227

Qy   644  CCCTATCACATCCATTCATTTGGTTCTCCGGATTCCAGGCA  680
          || .|||||||||:||:|:||||.:||:|||||||||||
Db   5228 CCUGUGCACUCCAUCCAUUUGGUUCACGGAUGCCUUUCA  5264

RESULT 10
US-10-134-288-1
;   Sequence 1, Application US/1013428
;   Publication No. US20030165520A1
;   GENERAL INFORMATION:
;   APPLICANT: Cao, Xuemei
;   APPLICANT: Sheppard, Mike
;   TITLE OF INVENTION: ATTENUATED FORMS OF BOVINE VIRAL DIARRHEA VIRUS
;   FILE REFERENCE: PC10435A
;   CURRENT APPLICATION NUMBER: US/10/134,288
;   CURRENT FILING DATE: 2002-04-29
;   PRIOR APPLICATION NUMBER: US/09/702,330
;   PRIOR FILING DATE: 2000-10-31
;   PRIOR APPLICATION NUMBER: 09/433,262
;   PRIOR FILING DATE: 1999-11-04
;   NUMBER OF SEQ ID NOS: 9
;   SOFTWARE: PatentIn Ver. 2.0
;   SEQ ID NO 1
;   LENGTH: 14078
;   TYPE: DNA
;   ORGANISM: Bovine Viral Diarrhea Virus
US-10-134-288-1

Query Match          21.9%;   Score 251.4;  DB 12;   Length 14078;
Best Local Similarity  94.2%;  Pred. No. 9.8e-71;
Matches  261;  Conservative  0;  Mismatches  16;  Indels  0;  Gaps  0;

Qy   404  ATGTGTAGCCGATGCCAAGGAAAGCATAGGAGGTTTGAAATGGACCGGGAACCTAAGAGT  463
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4532 ATGTGCAGCCGATGCCAGGGAAAGCATAGGAGGTTTGAAATGGACCGGGAACCTAAGAGT  4591

Qy   464  GCCAGATACTGTCTGAGTGTAATAGGCTGCATCCTGCTGAGGAAGGAGACTTTTGGGCA  523
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4592 GCCAGATACTGTCTGAGTGTAATAGGCTGCATCCTGCTGAGGAAGGTGACTTTTGGGCA  4651

Qy   524  GAGTCAAGCATGTTGGGCCTCAAGATCACCTACTTTGCACTGATGATGATGAAAGGTGAT  583
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4652 GAGTCGAGCATGTTGGGCCTCAAAATCACCTACTTTGCCCTGATGATGGATGAAAGGTGAT  4711

Qy   584  GACATCACAGAGTGGGCTGGATGCCAGCGTGTAGGTATCTCCCAGATACCCACAGAGTC  643
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db   4712 GATATCACAGAGTGGGCTGGATGCCAGCGTGTGGGAATCCCCAGATACCCACAGAGTC  4771

Qy   644  CCCTATCACATCCATTCATTTGGTTCTCCGGATTCCAGGCA  680
          |||||||||||||||:||:|:||||||:|||||||||
Db   4772 CCTTGTCACATCCATTCATTTGGTTCACGGATGCCTTTCA  4808

RESULT 11
US-10-228-406A-9
;   Sequence 9, Application US/10228406A
;   Publication No. US20030104612A1
;   GENERAL INFORMATION:
;   APPLICANT: Cao, Xuemei
;   APPLICANT: Zybarth, Gabriele
```

; TITLE OF INVENTION: GENERATION OF TYPE I/TYPE II HYBRID FORM OF BOVINE
; TITLE OF INVENTION: VIRAL DIARRHEA VIRUS FOR USE AS VACCINE
; FILE REFERENCE: PC11051A
; CURRENT APPLICATION NUMBER: US/10/228,406A
; CURRENT FILING DATE: 2002-08-27
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
;
; SEQ ID NO 9
; LENGTH: 16713
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: pNADL890 vector
US-10-228-406A-9

Query Match   21.9%;   Score 251.4; DB 14; Length 16713;
Best Local Similarity 94.2%;   Pred. No. 1.1e-70;
Matches 261; Conservative   0; Mismatches   16; Indels   0; Gaps   0;

Qy   404 ATGTGTAGCCGATGCCAAGGAAAGCATAGGAGGTTTGAAATGACCGGGAACCTAAGAGT 463
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4988 ATGTGCAGCCGATGCCAGGGAAAGCATAGGAGGTTTGAAATGGACCGGGAACCTAAGAGT 5047

Qy   464 GCCAGATACTGTGCTGAGTGTAATAGGCTGCATCCTGCTGAGGAAGGAGACTTTTGGGCA 523
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5048 GCCAGATACTGTGCTGAGTGTAATAGGCTGCATCCTGCTGAGGAAGGTGACTTTTGGGCA 5107

Qy   524 GAGTCAAGCATGTTGGGCCTCAGATCACCTACTTTGCACTGATGATGAAAGGTGTAT 583
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5108 GAGTCCGAGCATGTTGGGCCTCAAAATCACCTACTTTGCGCTGATGATGAAAGGTGTAT 5167

Qy   584 GACATCACAGAGTGGGCTGGACGCCAGCGTGTAGGTATCTCCCAGATACCCACAGAGTC 643
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5168 GATATCACAGAGTGGGCTGGACGCCAGCGTGTGGGAATCTCCCCAGATACCCACAGAGTC 5227

Qy   644 CCCTATCACATCTCATTTGGTTCTCGGATTCCAGGCA 680
          ||||||||||||||||||||||||||||||
Db   5228 CCTTGTCACATCTCATTTGGTTCACGGATGCCTTTCA 5264

RESULT 12
US-09-867-701-8500
; Sequence 8500, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSEQ for Windows Version 4.0
;
; SEQ ID NO 8500
; LENGTH: 339
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-8500

Query Match   20.5%;   Score 234.6; DB 10; Length 339;
Best Local Similarity 98.0%;   Pred. No. 4.4e-66;
Matches 248; Conservative   0; Mismatches   4; Indels   1; Gaps   1;

Qy   896 TTCCAACGTTGATGCCCCTTCTCTTTCCTCAAATCAATGTCAGGGAGTCAAAACGGGCTGT 955
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   23 TTTCAACGTTGATGCCCCTTCTCTTTCCTCAAATCAATGTCAGGGAGTCAAAAGGGGCTGT 82

Qy   956 AGCACAGGATGGAGTTTGATTTATCCCCTCTCCCCCAACACCTAGGAACTGAATCTTTTT 1015
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   83 AGCACAGGATGGGAGTTTGATTTTATCCTCCTCTCCCCCAACACCTAGGAACTGAATCTTTTT 142

Qy   1016 CTTTTTA-TTTTTTGAGATGGAGTCTTGCTCTGTTGCCCAGCTGGAGTGCAGTGGTGTGA 1074
          ||||||| ||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db   143 CTTTTTATTTTTTTTGAGATGGAGTCTCTGCTCTGTGTGCCCAGCTGGAGTGAAGTGGTGTGA 202

Qy   1075 TCTCAGCTTACTGCAACCCTCTGTCTCCCGGGTTCAAGCAATTCTCCCATCTCAGCCTCCT 1134
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   203 TCTAAGCTTACTGCAACCCTCTGTCTCCCAGGTTCAAGCAATTCTCCCATCTCAGCCTCCT 262

Qy   1135 GAGTAGCTGGGAT 1147
          |||||||||||||
Db   263 GAGTAGCTGGGAT 275

RESULT 13
US-09-814-353-4106/c
; Sequence 4106, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSEQ for Windows Version 4.0
;
; SEQ ID NO 4106
; LENGTH: 510
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 32
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-4106

Query Match   12.0%;   Score 137.6; DB 12; Length 510;
Best Local Similarity 96.6%;   Pred. No. 3.3e-34;
Matches 140; Conservative   0; Mismatches   5; Indels   0; Gaps   0;

Qy   1 GCCTTGGGTCAAGCAGAGAATATTAATAGGCAGGGGAATGCACCTGTAGCTAGTGGGCGCTA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   163 GCCTTGGGTCAAGCAGAGAATATTAATAGGCAGGGGAATGCACCTGTAGCTAGTGGGCGCTA 104

Qy   61 CTGCCAGCCTGAGAGAGGAGTGGCTGACTCTTGACCATGGCTGGGGTTCCTGAGGATGA 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   103 CTGCCAGCCTGAGAGAGGAGTGGCTGACTCTTGACCATGGCTGGGGTTCCTGAGGATGA 44

Qy   121 GCTAAACCCTTCCATGTACTGGGG 145
          ||||||||||||||||||||
Db   43 GCTAAACCCTTNCCATGTACCTCCG 19

RESULT 14
US-09-814-353-10412/c
; Sequence 10412, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:

; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10412
; LENGTH: 510
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 32
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-10412

Query Match   12.0%;   Score 137.6;   DB 12;   Length 510;
Best Local Similarity   96.6%;   Pred. No. 3.3e-34;
Matches 140; Conservative   0; Mismatches   5; Indels   0; Gaps   0;

Qy   1 GCCTTGGGTCAAGCAGAATATTAATAGGCAGGGGAATGCACCTGAGCTAGTGGGCGCTA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 163 GCCTTGGGTCAAGCAGAATATTAATAGGCAGGGGAATGCACCTGAGCTAGTGGGCGCTA 104

Qy   61 CTGCCAGCCTGAGAGGAAGTGGCTCGACTCTTGACCATGGCTGGGGTTCCTGAGGATGA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 103 CTGCCAGCCTGGAGAGGAAGTGGCTCGACTCTTGACCATGGCTGGGGTTCCTGAGGATGA 44

Qy 121 GCTAAACCCTTTCCATGTACTGGGG 145
          |||||||||| ||||||
Db   43 GCTAAACCCTTNCCATGTACCTCGG 19

RESULT 15
US-09-814-353-16796/c
; Sequence 16796, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661

; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16796
; LENGTH: 598
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-814-353-16796

Query Match   11.9%;   Score 137;   DB 12;   Length 598;
Best Local Similarity   96.6%;   Pred. No. 5.7e-34;
Matches 140; Conservative   0; Mismatches   5; Indels   0; Gaps   0;

Qy   1 GCCTTGGGTCAAGCAGAATATTAATAGGCAGGGGAATGCACCTGAGCTAGTGGGCGCTA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 234 GCCTTGGGTCAAGCAGAATATTAATAGGCAGGGGAATGCACCTGAGCTAGTGGGCGCTA 175

Qy   61 CTGCCAGCCTGAGAGGAAGTGGCTCGACTCTTGACCATGGCTGGGGTTCCTGAGGATGA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 174 CTGCCAGCCTGGAGAGGAAGTGGCTCGACTCTTGACCATGGCTGGGGTTCCTGAGGATGA 115

Qy 121 GCTAAACCCTTTCCATGTACTGGGG 145
          |||||||||| ||||||
Db 114 GCTAAACCCTTACCATGTACCTCGG 90

Search completed: December 1, 2003, 10:09:53
Job time : 1870 secs